TRAC-M-TM-15-032
September 2015

Validating the FOCUS Model Through an Analysis of Identity Fragmentation in Nigerian Social Media

TRADOC Analysis Center
700 Dyer Road
Monterey, CA 93943-0692

This study cost the
Department of Defense approximately
$61,000 expended by TRAC in
Fiscal Years 14-15.
Prepared on 20150922
TRAC Project Code # 060113

THIS PAGE INTENTIONALLY LEFT BLANK

TRAC-M-TM-15-032
September 2015

Validating the FOCUS Model Through an Analysis of Identity Fragmentation in Nigerian Social Media

Authors

MAJ Adam Haupt
Dr. Camber Warren

PREPARED BY:

ADAM HAUPT
MAJ, US Army
TRAC-MTRY

APPROVED BY:

CHRISTOPHER M. SMITH
LTC, US Army
Director, TRAC-MTRY

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE				*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** 24 September 2015	**3. REPORT TYPE AND DATES COVERED** Technical Memorandum, June 2014 to September 2015		
4. TITLE AND SUBTITLE Validating the FOCUS Model Through an Analysis of Identity Fragmentation in Nigerian Social Media				**5. PROJECT NUMBERS** TRAC Project Code 060113
6. AUTHOR(S) MAJ Haupt, Dr. Warren				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army TRADOC Analysis Center - Monterey 700 Dyer Road Monterey CA, 93943-0692				**8. PERFORMING ORGANIZATION REPORT NUMBER** TRAC-M-TM-15-032
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) Joint Warfare Analysis Center (JWAC) and TRADOC Analysis Center Headquarters (TRAC-HQ)				**10. SPONSORING/MONITORING AGENCY REPORT NUMBER** TRAC-M-TM-15-032
11. SUPPLEMENTARY NOTES Findings of this report are not to be construed as an official Department of the Army (DA) position unless so designated by other authorized documents.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited				**12b. DISTRIBUTION CODE**
13. ABSTRACT (maximum 200 words) Data-driven validation of the FOCUS model's capacity to predict the dynamics of social identity group (SIG) formation would allow the project to confirm the validity of the theoretical mechanisms encoded in the model. However, such efforts are currently inhibited by both an absence of high-resolution geo-spatially registered SIG data that could be systematically compared to the model's predictions. A spatial-temporal map of identity fragmentation in social media discourse would provide an ideal empirical target for FOCUS, allowing FOCUS to leap-frog competing projects which lack empirically validated, predictive capabilities, and consequently fail to satisfy the promise of generating believable probability distributions over the potential outcomes of operations intended to stabilize a region. This project, as a proof of concept, attempts to lay the foundation for the use of social media data to identify social variables that can be used to model acts of collective violence in Nigeria that can later be used to validate the FOCUS model.				
14. SUBJECT TERMS Big-Data, Social Media Analysis, Text Analytics, Statistical Modeling, Gaussian Kernel Density Interpolation, Nigeria, FOCUS				**15. NUMBER OF PAGES** 72
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified		**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

NOTICES

DISCLAIMER

Findings of this report are not to be construed as an official Department of the Army (DA) position unless so designated by other authorized documents.

REPRODUCTION

Reproduction of this document, in whole or part, is prohibited except by permission of the Director, TRAC, ATTN: ATRC, 255 Sedgwick Avenue, Fort Leavenworth, Kansas 66027-2345.

DISTRIBUTION STATEMENT

Approved for public release; distribution is unlimited.

DESTRUCTION NOTICE

When this report is no longer needed, DA organizations will destroy it according to procedures given in AR 380-5, DA Information Security Program. All others will return this report to Director, TRAC, ATTN: ATRC, 255 Sedgwick Avenue, Fort Leavenworth, Kansas 66027-2345.

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Data-driven validation of the FOCUS model's capacity to predict the dynamics of social identity group (SIG) formation would allow the project to confirm the validity of the theoretical mechanisms encoded in the model. However, such efforts are currently inhibited by both an absence of high-resolution geo-spatially registered SIG data that could be systematically compared to the model's predictions. A spatial-temporal map of identity fragmentation in social media discourse would provide an ideal empirical target for FOCUS, allowing FOCUS to leap-frog competing projects which lack empirically validated, predictive capabilities, and consequently fail to satisfy the promise of generating believable probability distributions over the potential outcomes of operations intended to stabilize a region. This project, as a proof of concept, attempts to lay the foundation for the use of social media data to identify social variables that can be used to model acts of collective violence in Nigeria that can later be used to validate the FOCUS model.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

CPU	Central Processing Unit
GPU	Graphic Processing Unit
FOCUS	Flow Of Communication Upon Society
JWAC	Joint Warfare Analysis Group
NPS	Naval Postgraduate School
RAM	Random Access Memory
ROM	Read Only Memory
SIG	Social Identity Group
TRAC	Training and Doctrine Command Analysis Center
TRADOC	Training and Doctrine Command

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to recognize Assistant Professor Dr. Camber Warren for his tireless efforts, ingenuity and initiative, which are solely responsible for the completion of this project.

THIS PAGE INTENTIONALLY LEFT BLANK

SECTION 1. INTRODUCTION

"Validating the FOCUS Model through an Analysis of Identity Fragmentation in Nigerian Social Media" is a project that was designed to gain valuable spatial-temporal data from social media sources. The results from this initial analysis is intended to eventually support the validation of the FOCUS model's capability to predict the dynamics of social identity groups (SIG) and separately predict violent conflict in a region. This document will discuss the steps that the study team took to process and analyze large volumes of social media data to gain statistically relevant insights into SIG and violent conflict.

1.1. BACKGROUND

The FOCUS model is a social dynamic model designed to predict changes in social identity groups over time and space. Data-driven validation of the FOCUS model's capacity to predict the dynamics of social identity group (SIG) formation would allow the project to confirm the validity of the theoretical mechanisms encoded in the model. However, such efforts are currently inhibited by both an absence of high-resolution geo-spatially registered SIG data that could be systematically compared to the model's predictions. A spatio-temporal map of identity fragmentation in social media discourse would provide an ideal empirical target for FOCUS, allowing FOCUS to leap-frog competing projects which lack empirically validated, predictive capabilities, and consequently fail to satisfy the promise of generating believable probability distributions over the potential outcomes of operations intended to stabilize a region. The Naval Postgraduate School and TRAC-MTRY researched the ability to generate these spatio-temporal maps and relevant real life data using Twitter data. This report discusses those efforts and the initial results which have the promise to support the validation of this social model.

1.2.1. Project History

In April of 2014 TRAC-MTRY had additional projects funds available for research. Dr. Camber Warren from the Defense Analysis department approached TRAC-MTRY with a desire to research the ability to use social media data to analyze social identity groups in different nations and the capability of social media data to predict violent conflict. TRAC-MTRY and JWAC decided to fund this project by funding the purchase of a 10% random sample of one

year's worth of worldwide Twitter data. The funds were transferred to NPS and Dr. Warren purchased the data through GNIP (Twitter data sales company) through NPS contracting. Though this project was projected to start in June 2014, the contracting process took much longer than anticipated. Twitter bought GNIP towards the end of the contracting process, which added additional months of contract negotiation. The purchased Twitter data was finally delivered in January 2015 and was the express property of NPS. Once NPS received the data Dr. Warren began organizing, processing and analyzing the data. By May, Dr. Warren had created the Python scrips to sort through the data. In August the analysis scripts were complete and Dr. Warren was able to generate informative heat maps of Twitter activity in both Nigeria and Syria and generated an academic paper that explained the process, methodology and results of his initial analytic efforts using Twitter social media. Though these product deliverables marked the end of this project, Dr. Warren is continuing to build on his initial successes and there is tremendous potential for follow on projects that will look to improve on the analytic methods used to gain greater understanding on the social dynamics of nations using social media.

1.2. FOCUS MODEL

The Flow of Communication Upon Society (FOCUS) model is an agent based social stability model designed by Dr. Steven Hall (NPS) and Dr. Ryan G. Baird (JWAC) that models the interaction of different population groups inside a network who are competing for political agendas and resources. It uses a geographically situated agent based modeling approach that changes over time. Agents, individually representing people and collectively representing a governable population, dialogue with various Factions, which are competing for the opportunity to establish governing policy. As time progresses in the model, loyalties across the population change and realign. FOCUS provides the means to explore the sensitivity of these emergent loyalties to the various modeled influences as well as to the influence of the geographical characteristics of the region, including population and media infrastructure distribution and density (Hall and Baird 2013, 2). Key to the validation of this model is the ability to use real world data as inputs into the model at time 0 and then compare the resulting modeling outputs at time *n* to the corresponding real world conditions at time *n*. Ultimately, it is the hope of this research team to build upon the methods employed in this research project to provide those time-step 0 inputs and the validating time-step *n* real world conditions.

1.3. PROBLEM STATEMENT

Can social media data be used to empirically validate the theoretical mechanisms encoded in the FOCUS model, by developing measures of identity fragmentation?

1.2.3. Issues for Analysis.

Issue 1: Can Social Media data provide relevant insight into a target country's social dynamic in time and space?

EEA 1.1.: Can social media data identify SIG?

EEA 1.2.: Can social media data identify or predict violent conflict?

1.4. CONSTRAINTS, LIMITATIONS AND ASSUMPTIONS.

Constraints limit the study team's options to conduct the study. *Limitations* are a study

team's inabilities to investigate issues within the sponsor's bounds. *Assumptions* are study-specific statements that are taken as true in the absence of facts.

- **Constraints**:
  - Complete by 30 September 2015.
  - Social Media data is limited to Twitter data from August 1st, 2013 to July 31st, 2014.
- **Limitations**:
  - Study is limited to the analysis of Nigeria and Syria in accordance with the approved study proposals.
  - Usable data was limited to geo-coded tweets which represented approximately 27% of the total data repository.
  - Key concepts and metrics were limited to social identity make-up, national identity, social unrest and violent conflict.

- **Assumptions**:
  - Nigeria and Syria provide a relevant test bed for developing theoretical metrics that will help provide insights into the SIGs and social unrest of all nations.
  - Geo-coded tweets provide sufficient representative data to produce relevant analysis on SIG and social unrest.

SECTION 2. METHODOLOGY

2.1. OVERVIEW

This section is meant to be a summary of the methodology employed in this project to gain insight into social identity groups and predict collective violence using social media. For greater detail into the processing and analysis of our archived twitter database refer to the attached technical paper written by Dr. Camber Warren entitled "Mapping the Rhetoric of Violence: Political Conflict Discourse and the Emergence of Identity Radicalization in Nigerian Social Media", which is located in Appendix A.

2.2. "BIG DATA"

The data for this research was an archived database of Twitter messages contracted through GNIP. The data represented a 10% random sample of all public messages sent through the Twitter network between 1 August 2013 and 31 July 2014. This archive constituted approximately 12 billion messages and in an uncompressed format was approximately 40 Terabytes. Although tweets are limited to 140 characters of content, the actual twitter file is considerably larger due to embedded metadata. An example of this additional metadata is user identification information, profile information and time and location information. As a part of the GNIP contract our twitter data was augmented with geo-location information in the form of longitude and latitude coordinates. However, roughly only 27% of the files had geo-location information. The implication of this was that only 27% of the data was useful for measuring spatial-temporal subjects from the corpus of information that we possessed (Warren 2015, 9). This usable dataset was further diminished when we began analysis of specific countries. In this project, the usable, geo-located dataset for Nigeria accounted for approximately 14 million tweets out of the 12 billion tweets in our archive.

2.3. HARDWARE CONFIGUATION

The sheer size of our archived Twitter database created tremendous challenges for storage and processing. Without sufficient storage and processing hardware the time it would take to process the 40 Terabytes information could take months of continuous run time. The data storage and processing tools that made this research feasible was a Central Processing Unit (CPU) / Graphic Processing Unit (GPU) hybrid server, designed to emphasize parallel computation and in-memory processing, which is crucial for largescale textual and geospatial analytics. The primary processors consisted of 4 x 12-core Intel Xeon E7-4860v2 CPUs for a total of 48 processing cores, which are capable of parallel processing. Additionally, there were two NVIDIA Tesla K40C GPU processors that equate to 5,760 GPU cores. GPUs have the unique ability to process numbers very quickly (millions of functions per second) and are crucial in high speed graphics and mathematical manipulations. The computer was further augmented with 64 x 32GB DDR3L server memory cards that provided the CPU/GPU with 2 Terabytes of Random Access Memory (RAM). This was perhaps the most critical component built into our CPU/GPU hybrid because it provided an enormous and efficient workbench for data processing. Finally, our CPU/GPU had 8 x 600GB SSD 6Gb/s SATA hard drives that equated to 4.8 terabytes of Read Only Memory (ROM) where the compressed Twitter data was archived. The combination of this hardware setup allowed for very rapid parallel processing that took advantage of very efficient parallel processors that could conduct all data manipulations on a RAM workbench that accelerated processing speeds.

It is worthy to note that initially we hoped to use the tremendous computational capabilities of the 5,760 GPU cores, but after significant research we discovered that GPUs were limited to mathematical number manipulation which is consistent with the needs of high speed computer graphics, but incompatible with textual analytics. Utilizing GPUs to process textual data is currently an important research topic in industry, but no actionable solutions are available at this time. The result of this discovery was that we were limited to the 48 CPU cores for processing data. Though this was less than what our team hoped, it still allowed us to process approximately 500,000 files per second, which equated to approximately seven hours of continuous run time to process the 12 billion files of Twitter data.

2.4. ANALYSIS METHODOLOGY

As a foundation for studying Nigeria, Dr. Warren examined three hypotheses to test the usefulness of social media data to predict regional violence (Warren 2015, 3):

- H1. Spatio-temporal regions with higher levels of violent political rhetoric will experience higher levels of violent political behavior.
- H2. Spatio-temporal regions with discourse characterized by more frequent reference to the country of “Nigeria” as a whole will experience lower frequencies of collective violence, because references to “Nigeria” indicate sentiment of national identity and belonging.
- H3. Spatio-temporal regions with discourse characterized by more frequent reference to the “Hausa” minority identity will experience higher frequencies of collective violence. This is based off of historical and social research that shows that the Housa minority group has been the primary center-of-gravity for violence in Nigeria.

In order to analyze these hypotheses we built a script in Python that would open each Twitter file and first see if it had a geo-coded location that was located in Nigeria and was regionally specific enough to show where in Nigeria the tweet occurred. These tweets were simultaneously being organized into 1-degree x 1-degree x 1-hour boxes of space-time along with the tweets’ content, stored entirely in RAM. These files were organized into a “key-value” store, which means that all records were indexed by a common key structure. The advantage of this setup is that it organizes all keys into a 'hash table', which allows for very fast record look-up speeds, even when the number of underlying records is very large (Warren 2015, 10). The result was 14,322,348 separate Twitter messages from inside Nigeria that were organized by space-time. This in-RAM dataset became the basis for our follow on analysis.

Next, three categories of searchable words were developed to answer our three hypotheses. Using the cross-language references in Wikipedia, different spelling variants of the conceptual category “Nigeria” (i.e. 'najeriya', 'naìjíríyà', 'naijiria') were identified and scripted into a hash table. This strategy was repeated for conceptual category “Housa.” Finally, a much more complex hash table was built for the concept of “armed conflict,” which included such words as

‘stabbing’, airstrike’, ‘soldier’, etc. In total a list of 366 English language terms, representing direct references to objects and actions associated with armed conflict, was developed to capture the concept of “armed conflict” in the content of our Nigerian dataset (Warren 2015, 33-39). These terms were then translated into the five most common non-English languages in Nigeria (i.e. French, Arabic, Hausa, Igo, and Yorbua), which yielded a total of 1,195 unique search strings.

With the search categories developed, each Twitter file in our Nigerian dataset was searched to identify matches to our search strings. Then we estimated a continuous spatial surface, representing the relative density of messages referencing each concept in a particular place and time using 2-dimensional binned Gaussian kernel density interpolation (Warren 2015, 14). Additionally, the same method was applied to the total Twitter message density to yield an estimated continuous spatial surface for the total Twitter message density. The final values developed were the estimated concept densities divided by the estimated total Twitter message densities over time and space. These four outputs could now be used as four distinct independent variables for statistical modeling. The sampling of the visual representation of these results can be viewed in Figure 1.

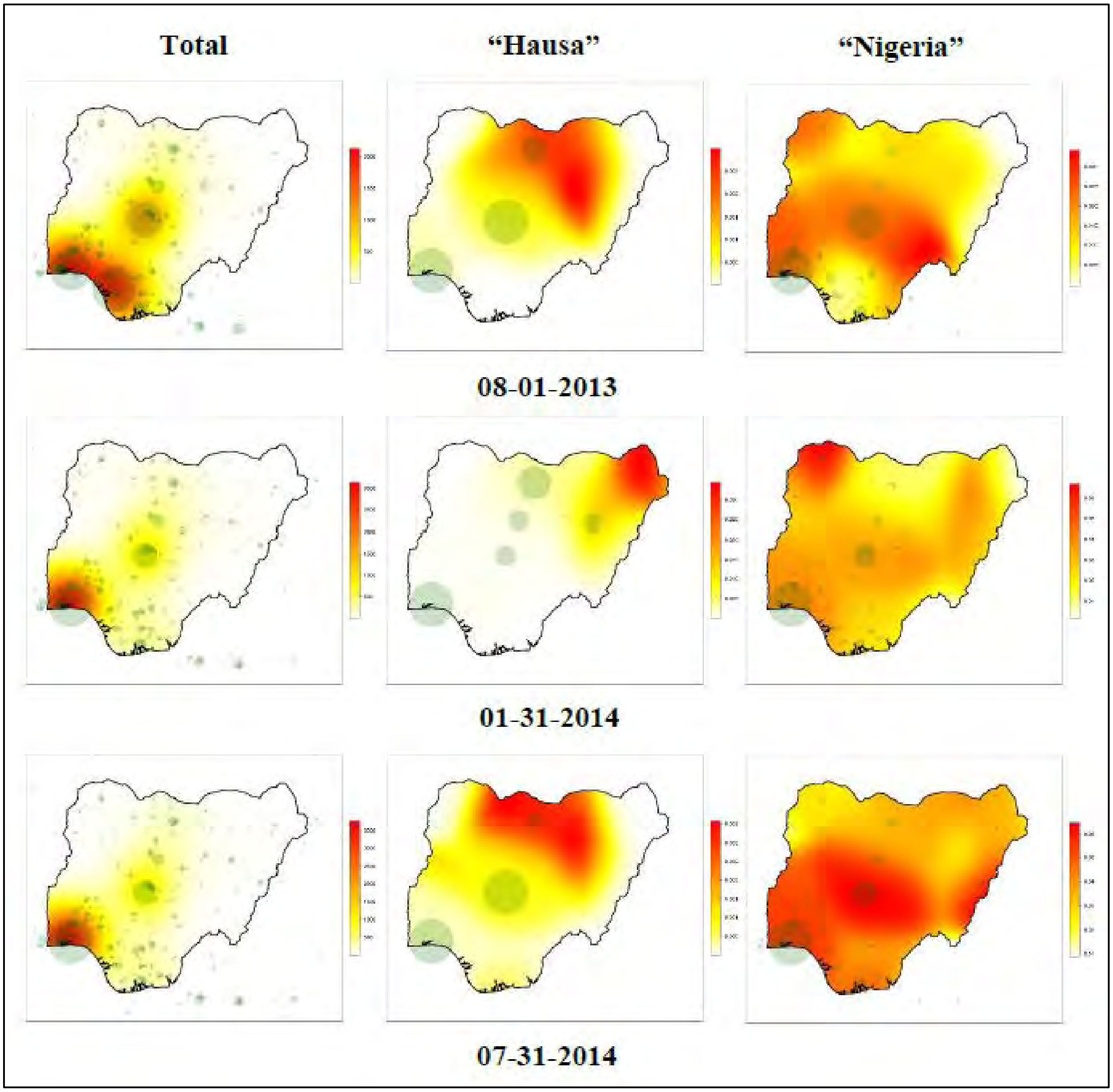

Figure 1. Spatio-Temporal Map of Nigeria (From Warren 2015, 18): These maps show the smoothed densities of the estimated 'total message density', the concept of 'Housa', and the concept of 'Nigeria' on three different days at the beginning, middle and end of our Nigerian dataset. Darker colors of red indicate higher densities of the concept; while lighter shades are lower densities (i.e. white is the most extreme low density). The green circles represent the actual Twitter message locations and the size of those circles represents comparative volume size.

In order to gain insight into the relevance of these variables to the modeling of violent conflict the team needed an accurate dataset of actual violent conflict of Nigeria that occurred during the span of our dataset. Using the Armed Conflict Location and Event Data Project

(ACLED) v5 database (Raleigh 2015), which contained a list of all armed conflict in Nigeria organized by date/time and latitude/longitude, we were able to populate a dependent variable that we could then use to build simple statistical models to test for the statistical significance of our three independent variables and answer our hypotheses (Warren 2015, 15).

SECTION 3. RESULTS

3.1. RESULTS OF ANALYSIS

To answer our three hypotheses, three models were developed using a heterogeneous point process model with a Strauss inter-point interaction function designed to flexibly capture patterns of spatial autocorrelation (Warren 2015, 15). The dependent variable for all three was acts of recorded violence in Nigeria. Model 1 used total Twitter message density as the sole independent variable. Model 2 included the four independent variables of 'Total Message Density', 'armed conflict', 'Nigeria', and 'Housa'. Model 3 included all of the before mentioned independent variables, but also included the interaction of 'Nigeria' x 'Hausa'. The results of these models showed several interesting facts. First, every independent variable was statistically significant with p-values less than 0.05. Second, all models were significant in their ability to model whether or not an act of collective violence in Nigeria occurred. Third, based off of the coefficients of each independent variable of Model 2, our hypotheses 1-3 were correct. Regions that experienced higher levels of violent rhetoric were more likely to experience collective violence; regions that had higher densities of identity to the national notion of 'Nigeria' experiences lower levels of collective violence, and regions that had higher reference densities to the concept of 'Housa' experienced higher levels of collective violence. Refer to *Figure 2* for a more in-depth look at the statistical modeling results.

We must note here that this analysis does not conclusively prove that social media data is 'reflective' or 'constructive' in nature, meaning we are not sure if social media discourse is just a reflection of events occurring (i.e. collective violence) or if it actually has a causation effect and leads to events occurring.

	Model 1	Model 2	Model 3
Total Density	8.1990 *** (1.1181)	13.5342 *** (2.7577)	25.6532 *** (3.2742)
"armed conflict"		3.2000 *** (0.3618)	3.3299 *** (0.3698)
"Nigeria"		-0.8351 ** (0.3105)	-4.4354 *** (0.5416)
"Hausa"		0.1518 *** (0.0246)	-1.2323 *** (0.1806)
"Nigeria" x "Hausa"			1.7509 *** (0.2226)
Intercept	1.6191 *** (0.1284)	-0.8584 * (0.3770)	2.0897 *** (0.5674)
Interpoint Interaction	0.0027 *** (0.0003)	0.0024 *** (0.0004)	0.0022 *** (0.0004)
AIC	-3882.43	-3917.65	-3984.23

Note: Coefficients from heterogeneous point process models. Standard error in parentheses. *p < 0.05, **p < 0.01, ***p < 0.001

Figure 2. Heterogeneous Point Process Modeling Results of three proposed models. Comparative ‘Goodness’ of the models was assessed using AIC. Significance of the independent variables was assessed by their associated p-values indicated by the asterisks.

3.2. DISCUSSION

The results of this research show that eventual application to the validation of the FOCUS model will be possible. Validating the FOCUS model requires the ability to map SIGs and sentiment concepts across space and time. By re-examining our primary ‘Issue for Analysis’ we can see that social media data does have the capability to inform on these inputs.

Issue for Analysis 1: Can Social Media Data provide relevant insight into a country's social dynamic in time and space? *Yes. In our proof-of-concept we were able to show that statistically relevant metrics could be identified over time and space that could partially identify SIGs and sentiment and could be used to model actual violent events in Nigeria.*

EEA 1.1: Can social media data identify SIGs? *Yes. Our metric of the concept "Housa" represented one of the most disenfranchised ethnic groups in Nigeria. The Housa are generally located in the north of the country, as seen in Figure 3. By examining the heat maps from Figure 1, we see that generally, the high density regions for the search concept of "Housa" were predominately clustered in the north of the country as well. Although this is not definitive proof that this SIG was identified, when coupled with the fact that this metric was statistically significant in predicting violent events in accordance with our hypothesis, we can make the case that social media data can identify SIGs.*

EEA 1.2: Can Social media data identify or predict violent conflict? *Yes. Our statistical modeling approach showed that we could create relevant statistical models that identified violent conflict by applying relevant independent variables pulled from social media data.*

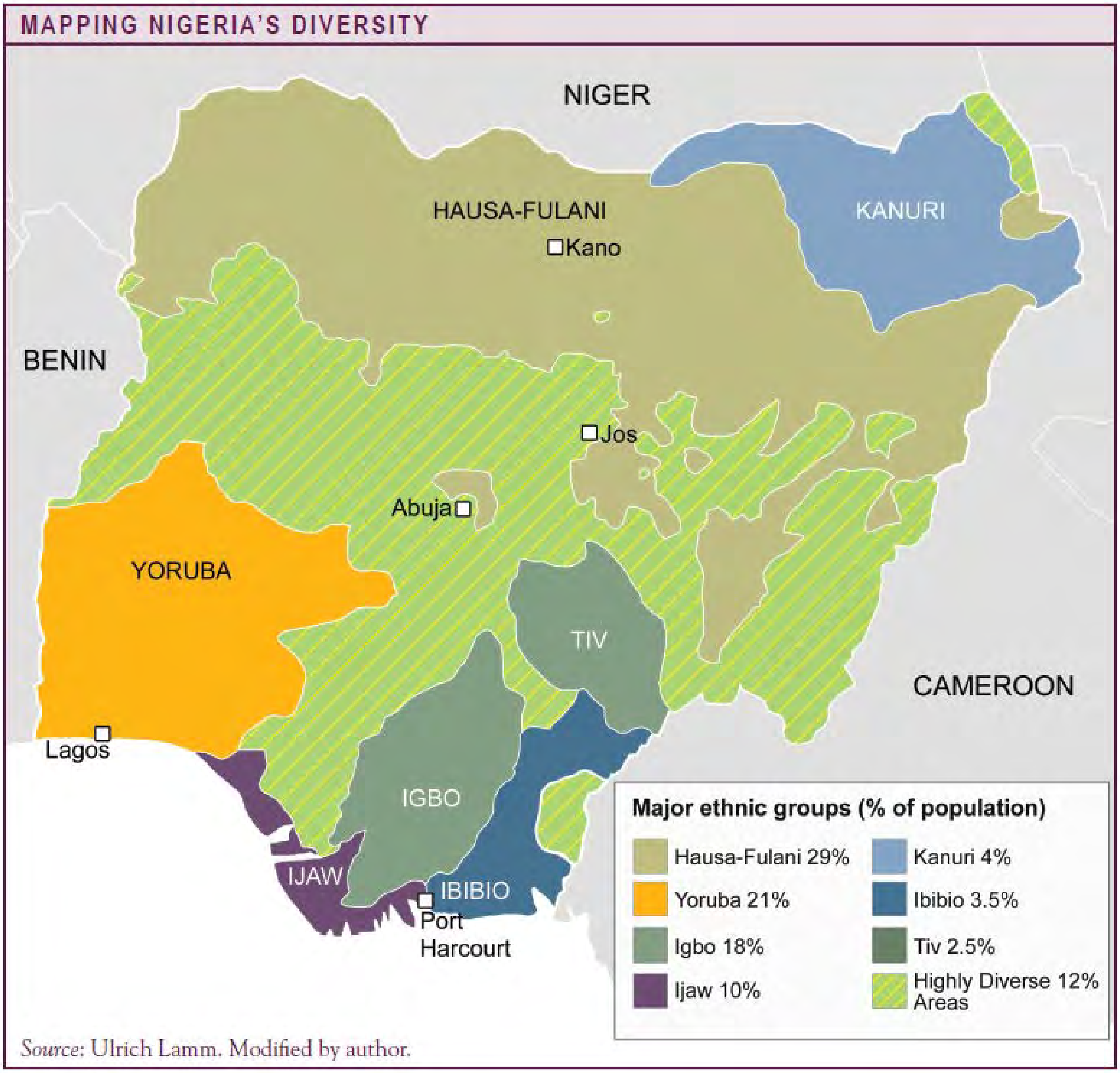

Figure 3. Ethnic Map of Nigeria (From Kwaja 2011, 3). This map shows that Housa ethnic group is primarily located in the northern regions of Nigeria, which is consistent with the high densities of the concept "Housa" seen in Figure 1.

SECTION 4. RECOMMENDATIONS

This research only represents the earliest phases of research designed to determine the feasibility of social media data use for measuring and modeling events occurring inside national borders. There is tremendous room for expanded research using the principals of spatial-temporal statistical analysis that this project explores. For a start we recommend exploring the scalability of applying social media data to regions of interest. Interesting results could be gained from more refined analysis of cities or districts within a country. Additionally, significant insights could be gained from enlarging the region of interest to multi-country regions and continents. Another important expansion of this research should address to which degree social media discourse is ‘reflective’ or ‘constructive’ in nature. One way to address this could be to model collective violence using social media variables in a time-series approach to see if social discourse can predict collective violence. Lastly, we would recommend that regional subject matter experts be used to create more refined and insightful search concepts that would better identify social identity groups and localized expressions that more effectively capture expressions of collective violence or socio-political conflict.

REFERENCES

Hall, Steven B., and Ryan G. Baird. *Modeling the Influence of Information Flow on Social Stability*. Monterey: Naval Postgraduate School, 2013.

Kwaja, Chris. *Nigeria's Pernicious Drivers of Ethno-Religious Conflict.* Washington, D.C.: The Africa Center For Strategic Studies, 2011.

Raleigh, Clionadh. 2015. http://www.acleddata.com (accessed August 1, 2015).

Warren, Camber. *Mapping the Rehtoric of Violence: Political Conflict Discourse and the Emergence of Identitiy Radicalization in Nigerian Social Media.* Monterey, CA: Department of Defense Analysis Naval Postgraduate School, 2015.

APPENDIX A. “MAPPING THE RHETORIC OF VIOLENCE: POLITICAL CONFLICT DISCOURSE AND THE EMERGENCE OF IDENTITY RADICALIZATION IN NIGERIAN SOCIAL MEDIA”

The attached academic paper, written by Assistant Professor Camber Warren, is the foundation for the content of this technical memo. It contains the technical solutions to the research problem that this project addressed and the methods and tools that were used to answer the elements of that problem.

Mapping the Rhetoric of Violence:

Political Conflict Discourse and the Emergence of Identity Radicalization in Nigerian Social Media[1]

T. Camber Warren

Department of Defense Analysis
Naval Postgraduate School

CamberW@gmail.com

Abstract

While there is widespread agreement amongst scholars and practitioners that processes of popular radicalization frequently underlie the generation of insurgent violence, an absence of high-resolution data has prevented existing work from directly validating this relationship. To begin to fill this gap, I seek to leverage new social media technologies to our advantage, by using them as a means of data collection. More specifically, I show that newly developed tools for geo-coding the sending locations of messages sent through the Twitter network, automated estimations of the sentiments expressed in those messages, and spatial interpolation of those estimates, can be used to generate dynamic, data-driven maps of national attachments and political extremism amongst the members of a given population. This approach is applied to the analysis of identity radicalization and fragmentation in Nigeria, over the period August 2013 to July 2014. The results demonstrate that network-analytic metrics derived from spatio-temporal variation in social media content hold substantial promise for enhancing our understanding of the conditions which most favor the emergence of political extremism and collective violence.

[1] Prepared for presentation at the Annual Meeting of the American Political Science Association, Sept. 3rd-6th, 2015, San Francisco, CA.

Introduction

A burgeoning body of literature increasingly points to the importance of communication dynamics in the generation of armed conflict and collective violence (Pierskalla and Hollenbach 2013; Shapiro and Weidmann 2015; Warren 2014, 2015; Weidmann 2015), and in particular the role played by polarization along newly politicized ethnic cleavages (Bhavnani and Miodownik 2009; Buhaug, Cederman, and Rød 2008; Cederman, Weidmann, and Gleditsch 2011; Cederman, Wimmer, and Min 2010). However, an absence of suitable data has prevented existing work from directly validating the relationship between patterns of political communication and patterns of political violence.

To begin to fill this gap, I seek to leverage new social media technologies to our advantage, by using them as a means of data collection. More specifically, I show that newly developed tools for geo-coding the sending locations of messages sent through the Twitter network, automated estimations of the sentiments expressed in those messages, and spatial interpolation of those estimates, can be used to generate dynamic, data-driven maps of national attachments and political extremism amongst the members of a given population.

As an initial plausibility probe, this approach is applied to the analysis of identity radicalization and fragmentation in Nigeria, over the period August 2013 to July 2014. In particular, I hypothesize that spatio-temporal variation in discursive references to particular conceptual categories will be systematically related to the generation of events of collective violence. Extending the argument presented in Warren (2014) and Warren (2015), I claim that this linkage represents a fundamental mechanism in the production of collective violence. In brief, large-scale violence requires the successful production and dissemination of political ideas justifying that violence. As a result, violence must be spoken into existence, before it can be

enacted. This implies that it may be possible to observe increases in the production of violent rhetoric prior to the emergence of violent acts, and perhaps even to use such measurements to predict the occurrence of collective violence before it erupts in actuality. Moreover, this perspective implies that variation in the basic conceptual categories of political communication could exercise profound effects on the likelihood of large-scale conflict. In regions where political discourse tends to deploy the unifying categories of "nation" and "country", it may be more difficult to generate the kinds of political ideation which justify violence against one's fellow citizens. In contrast, in regions where the dominant discourse revolves instead around narrow sectarian identities, it may be easier for political actors to generate the kinds of animosities that feed spirals of polarized violence. Nigeria provides a particularly interesting window on such dynamics, as the north of the country has recently been characterized by increasingly vociferous mobilization of the "Hausa" ethnic minority, by political actors seeking greater regional autonomy. I will thus examine the following hypotheses:

H1. Spatio-temporal regions with higher levels of violent political rhetoric will experience higher levels of violent political behavior.

H2. Spatio-temporal regions with discourse characterized by more frequent reference to the country of "Nigeria" as a whole will experience lower frequencies of collective violence.

H3. Spatio-temporal regions with discourse characterized by more frequent reference to the "Hausa" minority identity will experience higher frequencies of collective violence.

The Predictive Power of Social Media

With the surging global popularity of social media platforms, researchers from a variety of disciplines have begun seeking analytic approaches which might allow predictive insights to be derived from social media streams in an unsupervised fashion. While some have focused on the aggregate dynamics of popular culture (Agarwal, Xie, Vovsha, Rambow, et al. 2011; Asur and Huberman 2010; Bae and Lee 2012; Barbosa and Feng 2010; Benhardus and Kalita 2013; Bessi, Caldarelli, Vicario, Scala, et al. 2014; Cataldi, Caro, and Schifanella 2010; Golder and Macy 2011; Hansen, Arvidsson, Nielsen, Colleoni, et al. 2011; Jansen, Zhang, Sobel, and Chowdury 2009; Java, Song, Finin, and Tseng 2007; Kim, Bak, and Oh. 2012; Lerman and Ghosh 2010; Lerman and Hogg 2010; Leskovec, Adamic, and Huberman 2007; Lin, Keegan, Margolin, and Lazer 2014; Morris, Counts, Roseway, Hoff, et al. 2012; Naaman, Boase, and Lai 2010; Naveed, Gottron, Kunegis, and Alhadi 2011; Suh, Hong, Pirolli, and Chi 2010; Wu and Huberman 2007; Wu, Hofman, Mason, and Watts 2011), others have attempted to use metrics derived from individual messages to develop algorithms that 'learn' the underlying sentiments of individual communicators (Abbasi, Chen, and Salem 2008; Agarwal, Xie, Vovsha, Rambow, and Passonneau 2011; Bae and Lee 2012; Barbosa and Feng 2010; Bifet and Frank 2010; Bollen, Pepe, and Mao 2011; Dodds, Harris, Kloumann, Bliss, et al. 2011; Fan, Zhao, Chen, and Xu. 2014; Ghiassi, Skinner, and Zimbra 2013; Golder and Macy 2011; Huang, Peng, Li, and Lee 2013; Jiang, Yu, Zhou, Liu, et al. 2011; Mitchell, Frank, Harris, Dodds, et al. 2013; O'Connor,

Balasubramanyan, Routledge, and Smith 2010; Pak and Paroubek 2010; Stieglitz and Dang-Xuan 2012; Thelwall, Buckley, and Paltoglou 2011; Wang, Can, Kazemzadeh, Bar, et al. 2012). However, both approaches have face serious difficulties in the pursuit of systematic empirical validation. In particular, the lack of any systematic cross-linguistic and cross-cultural 'ground-truth' against which to compare automated sentiment classifications, has generally forced such researchers to limit themselves to single-language (usually English) texts drawn from limited domains (e.g. news reports, movie reviews, etc.).

In contrast, a more recent wave of scholarship has sought to develop metrics geared towards the generation of explicit predictions, which can be compared more directly to observed events. In particular, researchers have shown that mood-based signals drawn from aggregate streams of Twitter messages are partially predictive of swings in financial markets (Bollen, Mao, and Zeng 2011; Zhang, Fuehres, and Gloor 2011, 2012). Along similar lines, a number of researchers have found that political election results can be predicted with some accuracy through relatively simple counts of references to the opposing candidates (Adamic and Glance 2005; Bermingham and Smeaton 2011; Franch 2013; Gayo-Avello 2013; Lassen and Brown 2011; Metaxas and Mustafaraj 2012; Tumasjan, Sprenger, Sandner, and Welpe 2010; Wang, Can, Kazemzadeh, Bar, and Narayanan 2012). While such work has generated more convincing evidence that useful information can be derived from social media streams in an automated fashion, such 'predictions' have generally been limited to relatively simple outcomes, and have been somewhat limited in their ability to shed light on the actual mechanisms underlying the events of interest.

Taking a different angle on social media research, other researchers have sought to use these new communication media as sources of data on the behavior of underlying human

populations. Seen from this perspective, social media represent a new kind of human "macroscope", allowing researchers to measure quantities that would have previously remained opaque to observation, at a scale and resolution that would have previously been impossible to achieve. In this way, social media can serve as a new tool for developing enhanced understanding of the fundamental mechanisms underlying human social and political interactions. For instance, a number of works have begun investigating how cultural products achieve popularity, examining both the content-level and context-level factors that lead messages to be repeated, and developing new models of the dynamics of information diffusion (Aral and Walker 2012; Bakshy, Hofman, Mason, and Watts 2011; Bliss, Kloumann, Harris, Danforth, et al. 2012; Boyd, Golder, and Lotan 2010; Cha, Haddadi, Benevenuto, and Gummadi 2010; Dodds, Harris, Kloumann, Bliss, and Danforth 2011; Eisenstein, O'Connor, Smith, and Xing 2014; Golder and Yardi 2010; Golub and Jackson 2010; Gomez, Manuel, and Krause 2010; Hansen, Arvidsson, Nielsen, Colleoni, and Etter 2011; Kwak, Lee, Park, and Moon 2010; Pfitzner, Garas, and Schweitzer 2012; Romero, Meeder, and Kleinberg 2011; Shamma, Kennedy, and Churchill 2011; Stieglitz and Dang-Xuan 2012; Zaman, Herbrich, Gael, and Stern 2010). In a similar vein, researchers have begun to examine the forces underlying the generation of 'collective attention', combining empirical measures with simulation models of competition between 'memes', to examine the operation of ecological constraints on message reproduction (Benhardus and Kalita 2013; Cataldi, Caro, and Schifanella 2010; Hong and Davison 2010; Jungherr and Jurgens 2013; Lehmann, Gonçalves, Ramasco, and Cattuto 2012; Mehrotra, Sanner, Buntine, and Xie 2013; Mei, Liu, Su, and Zhai 2006; Sasahara, Hirata, Toyoda, Kitsuregawa, et al. 2013; Weng, Flammini, Vespignani, and Menczer 2012; Wu and Huberman 2007), while others have used data from social media streams to build models of the mechanisms

underlying the formation and dissolution of social ties between individuals (Bollen, Gonçalves, Ruan, and Mao 2011; Bond et al. 2012; Coviello et al. 2014; Fan, Zhao, Chen, and Xu. 2014; Frank, Mitchell, Dodds, and Danforth 2012; Golder and Yardi 2010; Gonzalez, Cuevas, Cuevas, and Guerrero 2011; Himelboim, McCreery, and Smith 2013; Kuehn, Martens, and Romero 2014; Lazer et al. 2009; Mitchell, Frank, Harris, Dodds, and Danforth 2013; Mutz 2002; Shalizi and Thomas 2011; Zamal, Faiyaz, and Ruths 2012)

Increasingly, such efforts are also being applied to the political domain, yielding substantial new insights into the dynamics of public opinion, electoral competition, and political persuasion (Adamic and Glance 2005; Ausserhofer and Maireder 2013; Barberá and Rivero 2014; Barberá 2014, 2015; Barberá, Jost, Nagler, Tucker, et al. 2015; Bermingham and Smeaton 2011; Bond and Messing 2015; Chadwick 2006, 2013; Conover et al. 2011; Conover, Gonçalves, Flammini, and Menczer 2012; Conover, Gonçalves, Ratkiewicz, Flammini, et al. 2011; DiGrazia, McKelvey, Bollen, and Rojas 2013; Farrell 2012; Feller, Kuhnert, Sprenger, and Welpe 2011; Golbeck and Hansen 2014; Grossman, Humphreys, and Sacramone-Lutz 2014; Himelboim, McCreery, and Smith 2013; Lawrence, Sides, and Farrell 2010; Monroe, Colaresi, and Quinn 2008; Mustafaraj, Finn, Whitlock, and Metaxas 2011, 2011; Parmelee and Bichard 2012; Prior 2007; Ringsquandl and Petkovic 2013; Shirky 2011; Stieglitz and Dang-Xuan 2012; Wojcieszak and Mutz 2009; Yardi and Boyd 2010). In addition to the study of 'normal' politics, researchers are also increasingly using metrics derived from social media to shed new light on the dynamics of social mobilization, political polarization, and collective violence (Aday et al. 2010; Bailard 2015; Brandt, Freeman, and Schrodt 2011, 2014; Colbaugh and Glass 2012; Conover et al. 2013; Gleason 2013; Gohdes 2015; Hammond and Weidmann 2014; Howard and Hussain 2013, 2011; Hussain and Howard 2013; Lotan, Graeff, Ananny, Gaffney, et al. 2011;

Martin-Shields and Stones 2014; Metternich, Dorff, Gallop, Weschle, et al. 2013; Metzger et al. 2014; Munger 2014; Pierskalla and Hollenbach 2013; Ramakrishnan et al. 2014; Ritter and Trechsel 2014; Schroeder, Everton, and Shepherd 2014; Shapiro and Weidmann 2015; Siegel 2014; Theocharis 2013; Tudoroiu 2014; Tufekci and Wilson 2012; Wang, Gerber, and Brown 2012; Ward et al. 2013; Warren 2015; Windt and Humphreys 2014; Wolfsfeld, Segev, and Sheafer 2013; Zeitzoff, Kelly, and Lotan 2015; Zeitzoff 2013). Moreover, while such research has generally found that such technologies decrease stability in weak-state environments, other researchers have pointed to the ability of authoritarian governments to also turn such tools to their advantage (Gohdes 2015; Howard, Agarwal, and Hussain 2011; Kalathil and Boas 2003; King, Pan, and Roberts 2013; Lynch 2011; Morozov 2011; Munger 2014; Rød and Weidmann 2015).

A Spatio-Temporal Approach

In most of the analyses reported above, metrics were calculated based on units of analysis characterized by individual users, or individual messages. The difficulty with such approaches, when attempting to make statistical judgements concerning the underlying population, is that the sample is likely to be strongly biased along a number of dimensions. It is well known that use of social media correlates with a number of demographic characteristics, including age and wealth, and that social media users are therefore unlikely to provide a fully representative sample of the underlying population (Ansolabehere and Hersh 2012; Barberá and Rivero 2014; Mislove, Lehmann, Ahn, Onnela, et al. 2011). As a result, metrics for which "users" are in the denominator (i.e. positive messages per user per day) are likely to be similarly biased.

The approach adopted here is instead to characterize the relevant metrics as functions of space-time units, rather than as proportions of users. Here, I take inspiration from recent work which has shown improvements in our abilities to make automatic judgements of geographic location from unstructured text in Twitter user profiles (Blanford, Huang, Savelyev, and MacEachren 2015; Cheng, Caverlee, and Lee 2010; Compton, Jurgens, and Allen 2014; Conover et al. 2013; Hawelka et al. 2014; Kaltenbrunner et al. 2012; Kulshrestha, Kooti, Nikravesh, and Kp. 2012; Lee and Sumiya 2010; Leetaru, Wang, Cao, Padmanabhan, et al. 2013; Mitchell, Frank, Harris, Dodds, and Danforth 2013; Nemeth, Mauslein, and Stapley 2014; Takhteyev, Gruzd, and Wellman 2012; Yuan, Cong, Ma, Sun, et al. 2013). This approach allows researchers to greatly expand the sample of Twitter messages which can be geo-referenced (from around 2% to 27%), by avoiding the need for GPS coordinates, and instead relying on the user-reported hometowns from their public profiles.

The starting point for this analysis is an archived database of Twitter messages, representing a fully randomized 10% sample of all public messages sent through the Twitter network between August 1st, 2013 and July 31st, 2014; approximately 12 billion messages in total.[2] In uncompressed format, this archive represents approximately 40 Terabytes of textual data, and so the very scale which offers this new "macroscope" also represents a challenge for standard computational approaches, which search across strings in serial order. The solution adopted here is to script the production of "in-memory" database indexes, organized to reflect bins of space, time, and other nested concepts. In particular, I utilize what is known as a "key-value" store, which means that all records are indexed by a common key structure, which is just

[2] Archive licensed through agreement between Twitter, Inc. and the U.S. Naval Postgraduate School, as part of the "Global Data Initiative." See www.camberwarren.net/gdi.

a string describing membership in some set of containers in which many individual records are stored. The database is a modified version of the open-source Aerospike database,[3] which I have expanded to allow for highly-parallelized loading of data into RAM, by creating separately threaded communication channels for each logical CPU core in the system, allowing 'swarms' of parallel computational workers to operate in tandem, and avoid resource conflicts, without the need for hierarchical control structures. The advantage of this setup is that it organizes all keys into a 'hash table', which allows for very fast record look-up speeds, even when the number of underlying records is very large.

Our first task is to use this memory structure to reference each message to a location in space, given by latitude and longitude coordinates. To do so, I draw on data from the geonames.org gazetteer, an open-source database of named geographic places. The database contains references to over 10 million individual locations, with latitude and longitude coordinates, in addition to over 2 million alternate names and spellings, spanning over a 100 languages. Converting this information into a searchable form requires first 'tokenizing' the individual strings into meaningful chunks (i.e. words and phrases). This process is relatively straightforward for English, as it makes consistent use of spaces to differentiate words. However, this pattern is far from universal in other languages. For instance, ideographic languages such as Chinese and Japanese generally use long strings of characters with no spaces in between words, while Vietnamese uses spaces in between each syllable of a single word. Moreover, sometimes atomic concepts, such as "China", are represented by 'words' composed of

---

[3] See http://www.aerospike.com/. The database application also makes use of a modified version of the UltraJSON python library (https://github.com/esnme/ultrajson), which I have expanded to allow for bulk parsing of large, multiline text files, and a modified version of the RE2 python library (https://github.com/facebook/pyre2/), expanded to allow for grouped regular expression pattern matching using hierarchically nested terms. All modified source code will be redistributed on an open-source basis. Contact author for details.

one, two, three, or more ideographic characters. In Cambodian, a number of common place names require as many as eight ideographic word-characters to write the string referring to a single city. Thus, the very notion of what counts as a “word” or “phrase” is difficult to generalize across languages. The solution generally adopted in the works cited above, has been to either ignore the problem by focusing on English place names, or to develop language-specific parsers for particular applications. But this requires expensive computations, as each parser must actually read and make sense of the string in order to determine the proper word/phrase boundaries, and so cannot be feasibly implemented for search across a large number of languages simultaneously.

Instead, I construct a generic multilingual phrase index by segmenting each text string arbitrarily, without expending any effort to ‘read’ or make semantic sense of the underlying text. To do so, I make use of a particular text encoding format, known at “UTF-8”, which has the advantage of coding all characters in fixed-size arrays of bytes. A roman letter, such as “a” for instance, is stored in a single byte, whereas nearly all ideographic characters in common use are stored as either 3 bytes or 6 bytes. This means that whereas roman scripts can be split into words by breaking at every space, ideographic scrips can be broken into potential words by splitting the string in byte lengths of multiples of 3. Some of the resulting sub-strings will be nonsense, but they can be easily screened out by attempting to re-encode the bytes as valid UTF-8 characters, and discarding any uninterpretable sub-strings. Arbitrary phrases are thus constructed from each string by first splitting at every space, and then taking any remaining non-roman characters and extracting all unique substrings with lengths equal to multiples of 3, and then concatenating the resulting words into space-separated sequences (i.e. ‘phrases’) consisting of all unique sub-sequences with length less than some maximum phrase length. I in the results reported below, I

allow for phrase lengths up to 9 ‘words’, to account for difficult strings such as “*Cộng hòa Xã hội chủ nghĩa Việt Nam*”, which is the name of the country of “Vietnam” written in Vietnamese, and “ក្រុងភ្នំពេញ”, which is the name of the city of “Phnom Penh” written in Cambodian. Each of these phrases is then separately indexed in an in-memory hash table, as described above. The result is a search index composed of approximately 23 million unique text phrases.

Input search strings are taken from the “Location” field associated with each Twitter message, which is simply a box into which users can type free-form descriptions of the location (usually a hometown) from which they are sending their messages. These input strings are tokenized through the same procedure, allowing one-to-one matching of exact phrases. When multiple matching strings are found, the algorithm narrows the potential matches by first checking for nested overlaps between administrative units, such as “Ohio”, and specific places, such as “Springfield”, and then prioritizes matches to more specific places over matches to more general areas. To break further ties, the algorithm then relies on a simple measure of the “salience” of the information in the search result, by assigning a score to each potential match, given by:

$$S_i = \left(\sqrt{P+1}\right)(L^2)$$

where P is the total population of the place, as recorded in the Geonames database, and L is the byte-length of the matching character string.

For each record, we first check whether GPS coordinates are available (less than 2% of the sample), and if they are not then we attempt to match any location text using the procedure described above. Records for which no matching locations can be found, or which can only be matched at level of countries or top-level administrative units, are discarded. The remaining records (approximately 27% of the original sample) are then parsed, assigned latitude, longitude,

and timestamp coordinates, and stored in a separate key-value database, in which the keys are given by unique combinations of discrete units of space and time. In this way, the keys of the database function as spatio-temporal indexes, allowing for high-speed access of chunks of records defined by discrete ranges of time and space. The chunks are defined in units of latitude/longitude degrees and hours, so that each storage bin holds the records for a 1-degree x 1-degree x 1-hour box of space-time. The result is an in-memory structured representation of each record, stored entirely in RAM, recording the full text of each message, the estimated geo-coordinates of the user's sending location, and the date and time when the message was sent.

Using this approach, I identify 14,322,348 separate Twitter messages sent from within the boundaries of Nigeria, between August 1st, 2013 and July 31st, 2014. This set of records forms the basis for the results reported below. In order provide predictive leverage on the location and timing of violent events, I seek to side-step the thorny issues associated with cross-cultural interpretations of complex symbols, attitudes, and sentiments, and focus instead on discursive references to particular “concepts”, for which more rigorous bounds can be defined on a cross-cultural basis. In particular, I aim to capture simple indicators of three concepts, with differing levels of complexity: (1) a country (“Nigeria”) understood a fixed referent by those familiar with the term, (2) a group (“Hausa”) representing a locus of recent political struggle, and (3) a category of action (“armed conflict”) which can be objectively defined but which is described in practice through a wide array of terms.

The concepts of “Nigeria” and “Hausa”, while complex in a sociological sense, are relatively easy to search for in text form. Even across the major linguistic communities in Nigeria, these terms tend to be spelled in approximately the same way. Using the cross-language references in Wikipedia, I identify seven local spelling variants for “Nigeria” ('nijeriya',

'najeriya', 'naìjíríyà', 'naijiria', 'naigeria', 'nàìjíríà', and 'naijiriya') and four local spelling variants for "Hausa" ('bahaushe', 'bahaushiya, 'hausawa', and 'haoussa').

The concept of "armed conflict", in contrast, represents a more difficult search task, as it can be referenced through a wide variety of specific objects and actions (e.g. 'stabbing', 'airstrike', 'soldier', etc.), all of which need to be jointly recognized as members of the overarching concept. To accomplish this on a cross-linguistic basis, I first cross reference existing lexicons (Harvard Inquirer, MPQA) to develop a list of 366 English language terms representing direct references to objects and actions associated with armed conflict (see Appendix), taking care to include all forms of relevant nouns and verbs. I then use scripted access to the Google Translate API (https://translate.google.com/) to attempt to translate each term into the five most common non-English languages in Nigeria: French, Arabic, Hausa, Igo, and Yorbua. The results of this machine translation exercise are shown in Table A1, with blank cells indicating either that no translation was possible or that the original term was selected as the best translation. As can clearly be seen, the French and Arabic translations achieve more thorough coverage than the smaller Nigerian languages, but there is good general coverage across all concepts and languages. Collapsing this table into a searchable index yields 1,195 unique search strings, which are stored and indexed in a separate database using the tokenization procedures described above.

For each concept and each day, I estimate a continuous spatial surface, representing the relative density of messages referencing that concept in a particular place and time. The smoothing is conducted using 2-dimensional binned Gaussian kernel density interpolation. For each concept, for each day I estimate a separate smoothed density, treating as separate points each message containing the concept, and then calculate a separate smoothed density surface

using the full sample of messages, regardless of content. The final values reported for each concept are then the concept density estimated at a given location in space-time, divided by the total estimated message density at that location. The result is a smooth surface estimating the likelihood that a given location will produce a token of a given concept, relative to the total volume of tokens produced at that location.

Figure 1 shows a color-scale representation of the smoothed densities of total message volume and the relative densities of the concepts of "Nigeria" and "Hausa", on days at the beginning, middle, and end of our period of study, with red indicating higher levels and yellow indicating lower levels. The green circles show the actual locations of the messages used to produce the smoothed surfaces, with larger bubbles representing a greater volume of messages. As can clearly be seen, these metrics generate substantial content-based variation which is not simply reflective of the underlying volume of messages. Moreover, the geographic distribution of references to these terms varies significantly, with references to "Hausa" occurring much more frequently in the north of the country where Hausa communities represent a larger proportion of the population.

Statistical Models and Results

In order to draw inferences regarding the relationship between these metrics and the emergence of collective violence, I estimate heterogeneous point process models with a Strauss inter-point interaction function designed to flexibly capture patterns of spatial autocorrelation without forcing the analyst to pre-specify spatial units at any particular resolution (see Warren (2015) for a discussion). The dependent variable is measured using the ACLED v5 database, from which I build a list of the locations of all violent armed conflict events occurring within

Nigeria, from September 1st, 2013 to July 31st, 2014 (n = 1,427). For each event, covariate values are associated with the event by taking the daily smoothed surfaces described above and averaging across a temporal window stretching back over the previous 30 days. Randomly generated control points generated for statistical inference are spread evenly within this space-time box. Regression modelling then proceeds by comparing the covariate distributions observed at the random controls points, to the covariates observed at the actual event locations.

The results are presented in Table 1. Model 1 is a baseline specification which includes only total message density and the interpoint interaction function. Model 2 adds in the covariate surfaces capturing the relative density of our concepts, “armed conflict”, “Nigeria”, and “Hausa.” Finally, Model 3 add an interaction terms between “Nigeria” and “Hausa.” Taken as a whole, the results demonstrate that substantial predictive leverage can be gained through metrics derived from the content of social media messages. Comparing Model 1 to Model 2, we can that the AIC score improves with the addition of our content-based metrics, indicating that the results are not driven simply by differences in the penetration of the medium in different areas of the country. Rather, it appears that variation in the content of the messages provides additional predictive leverage over the likely locations of armed conflict events. In particular, the positive and statistically significant ($p < 0.001$) coefficient for “armed conflict” indicates that areas where people speak with more violent discourse are also areas that are more likely to generate actual events of violence. Moreover, the negative and significant coefficient for “Nigeria” ($p < 0.01$) indicates that areas where people make more frequent references to the country as a whole are less likely to generate internal collective violence. In contrast, the positive and significant coefficient for “Hausa” ($p < 0.001$) indicates that discursive references to this polarizing sectarian identity are systematically associated with higher levels of actual violence. Finally, the

positive and significant results for the interaction term between "Nigeria" and "Hausa" ($p < 0.001$) indicates that the most violent-prone configuration of these variables occurs in areas where "Nigeria" and "Hausa" are referenced with high joint density.

Conclusion

The results presented here thus provide new evidence for the importance of communication dynamics in the production of collective violence. Moreover, they demonstrate that it is possible, even with very simple metrics, to begin to differentiate forms of collective discourse that are more prone to be associated with actual events of collective violence. In particular, the evidence presented here indicates that discourses revolving around integrative national identities are likely to be less prone to the generation of collective violence than discourses that focus on divisive sectarian identities, while also pointing to the possibility that it is actually the confluence of these categories that is most strongly associated with the production of violence.

However, based on the very preliminary results presented here, a number of questions remain. While these associations generate substantial predictive leverage, it is not clear whether they arise due to "reflective" mechanisms, through which discourse comes to mirror existing events on the ground, or due to "constructive" mechanisms, through which discourse produces events that would not otherwise have occurred. Moving forward, closer attention to the temporal dynamics underlying these processes may make it possible to begin to disentangle the direction of these causal arrows.

Figure 1. Relative Spatio-Temporal Density of Discursive Concepts

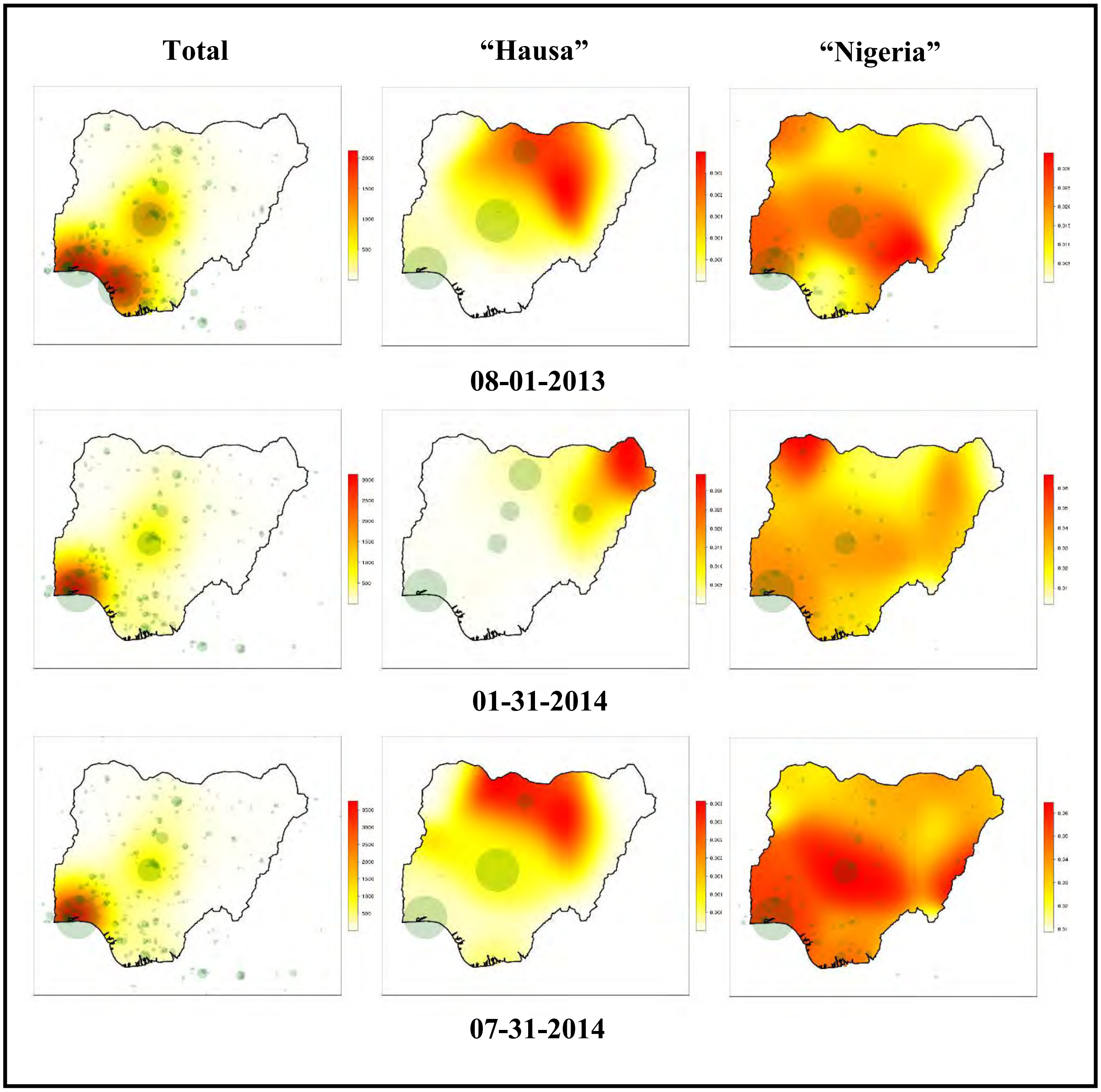

Table 1. Point Process Models of Violent Event Locations

	Model 1		Model 2		Model 3	
Total Density	8.1990	***	13.5342	***	25.6532	***
	(1.1181)		(2.7577)		(3.2742)	
"armed conflict"			3.2000	***	3.3299	***
			(0.3618)		(0.3698)	
"Nigeria"			-0.8351	**	-4.4354	***
			(0.3105)		(0.5416)	
"Hausa"			0.1518	***	-1.2323	***
			(0.0246)		(0.1806)	
"Nigeria" x "Hausa"					1.7509	***
					(0.2226)	
Intercept	1.6191	***	-0.8584	*	2.0897	***
	(0.1284)		(0.3770)		(0.5674)	
Interpoint Interaction	0.0027	***	0.0024	***	0.0022	***
	(0.0003)		(0.0004)		(0.0004)	
AIC	-3882.43		-3917.65		-3984.23	

Note: Coefficients from heterogeneous point process models. Standard error in parentheses.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

References

Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. 2008. "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums." *ACM Transactions on Information Systems (TOIS)* 26(3): 12.

Adamic, Lada A., and Natalie Glance. 2005. "The political blogosphere and the 2004 US election: divided they blog." In *Proceedings of the 3rd international workshop on Link discovery*, ACM, p. 36–43.

Aday, Sean et al. 2010. "Blogs and bullets: New media in contentious politics." *Report no.* 65.

Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. "Sentiment analysis of twitter data." In *Proceedings of the Workshop on Languages in Social Media*, Association for Computational Linguistics, p. 30–38.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate." *Political Analysis* 20(4): 437–459.

Aral, Sinan, and Dylan Walker. 2012. "Identifying influential and susceptible members of social networks." *Science* 337(6092): 337–341.

Asur, Sitaram, and Bernardo Huberman. 2010. "Predicting the future with social media." *International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) IEEE/WIC/ACM* 1: 492–499.

Ausserhofer, Julian, and Axel Maireder. 2013. "National politics on Twitter: Structures and topics of a networked public sphere." *Information, Communication & Society* 16(3): 291–314.

Bae, Younggue, and Hongchul Lee. 2012. "Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular twitterers." *Journal of the American Society for Information Science and Technology* 63(12): 2521–2535.

Bailard, Catie Snow. 2015. "Ethnic conflict goes mobile Mobile technology's effect on the opportunities and motivations for violent collective action." *Journal of Peace Research*.

Bakshy, Eytan, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. "Everyone's an influencer: quantifying influence on twitter." In *Proceedings of the fourth ACM international conference on Web search and data mining*, ACM, p. 65–74.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23(1): 76–91.

Barberá, Pablo. 2014. "How Social Media Reduces Mass Political Polarization: Evidence from Germany, Spain, and the U.S."

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, et al. 2015. "Tweeting From Left to Right Is Online Political Communication More Than an Echo Chamber?" *Psychological Science*.

Barberá, Pablo, and Gonzalo Rivero. 2014. "Understanding the political representativeness of Twitter users." *Social Science Computer Review*.

Barbosa, Luciano, and Junlan Feng. 2010. "Robust sentiment detection on twitter from biased and noisy data." In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Association for Computational Linguistics, p. 36–44.

Benhardus, James, and Jugal Kalita. 2013. "Streaming trend detection in twitter." *International Journal of Web Based Communities* 9(1): 122–139.

Bermingham, Adam, and Alan F. Smeaton. 2011. "On using Twitter to monitor political sentiment and predict election results." In *Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP)*, Dublin City University.

Bessi, Alessandro, Guido Caldarelli, Michela Del Vicario, Antonio Scala, et al. 2014. "Social determinants of content selection in the age of (mis) information." In *Social Informatics*, Springer International Publishing, p. 259–268.

Bhavnani, Ravi, and Dan Miodownik. 2009. "Ethnic polarization, ethnic salience, and civil war." *Journal of Conflict Resolution* 53(1): 30–49.

Bifet, Albert, and Eibe Frank. 2010. "Sentiment knowledge discovery in Twitter streaming data." In *Proceeding of 13th international conference on Discovery Science Conference*, Berlin Heidelberg: Springer, p. 1–15.

Blanford, Justine I., Zhuojie Huang, Alexander Savelyev, and Alan M. MacEachren. 2015. "Geo-Located Tweets. Enhancing Mobility Maps and Capturing Cross-Border Movement." *PloS one* 10(6).

Bliss, Catherine A., Isabel M. Kloumann, Kameron Decker Harris, Christopher M. Danforth, et al. 2012. "Twitter reciprocal reply networks exhibit assortativity with respect to happiness." *Journal of Computational Science* 3(5): 388–397.

Bollen, Johan, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011. "Happiness is assortative in online social networks." *Artificial life* 17(3): 237–251.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market." *Journal of Computational Science* 2(1): 1–8.

Bollen, Johan, Alberto Pepe, and Huina Mao. 2011. “Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.” In *Proceedings of the fifth international aaai conference on weblogs and social media (ICWSM 2011)*, ed. Barcelona July. Spain, p. 1–10.

Bond, Robert M. et al. 2012. “A 61-million-person experiment in social influence and political mobilization.” *Nature* 489(7415): 295–298.

Bond, Robert M., and Solomon Messing. 2015. “Quantifying Social Media’s Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook.” *American Political Science Review* 109(1): 62–78.

Boyd, Danah, Scott Golder, and Gilad Lotan. 2010. “Tweet, tweet, retweet: Conversational aspects of retweeting on twitter.” *Hawaii International Conference on System Sciences (HICSS)* (43): 1–10.

Brandt, Patrick T., John R. Freeman, and Philip A. Schrodt. 2014. “Evaluating forecasts of political conflict dynamics.” *International Journal of Forecasting* 30(4): 944–962.

Brandt, Patrick T., John R. Freeman, and Philip A. Schrodt. 2011. “Real time, time series forecasting of inter-and intra-state political conflict.” *Conflict Management and Peace Science* 28(1): 41–64.

Buhaug, Halvard, Lars-Erik Cederman, and Jan Ketil Rød. 2008. “Disaggregating ethno-nationalist civil wars: A dyadic test of exclusion theory.” *International Organization* 62(03): 531–551.

Cataldi, Mario, Luigi Di Caro, and Claudio Schifanella. 2010. “Emerging topic detection on twitter based on temporal and social terms evaluation.” In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, p. 4. ACM.

Cederman, Lars-Erik, Nils B Weidmann, and Kristian Skrede Gleditsch. 2011. “Horizontal inequalities and ethnonationalist civil war: A global comparison.” *American Political Science Review* 105(03): 478–495.

Cederman, Lars-Erik, Andreas Wimmer, and Brian Min. 2010. “Why do ethnic groups rebel? New data and analysis.” *World Politics* 62(01): 87–119.

Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and P. Krishna Gummadi. 2010. “Measuring User Influence in Twitter: The Million Follower Fallacy.” *ICWSM* 10(17): 30.

Chadwick, Andrew. 2006. *Internet Politics: States, Citizens, and New Communication Technologies*. Oxford, UK: Oxford University Press.

Chadwick, Andrew. 2013. *The hybrid media system: politics and power*. Oxford, UK: Oxford University Press.

Cheng, Zhiyuan, James Caverlee, and Kyumin Lee. 2010. “You are where you tweet: a content-based approach to geo-locating twitter users.” In *Proceedings of the 19th ACM international conference on Information and knowledge management*, ACM, p. 759–768.

Colbaugh, Richard, and Kristin Glass. 2012. “Early warning analysis for social diffusion events.” *Security Informatics* 1(1): 1–26.

Compton, Ryan, David Jurgens, and David Allen. 2014. “Geotagging one hundred million twitter accounts with total variation minimization.” *IEEE International Conference on Big Data (Big Data)*: 393–401.

Conover, Michael, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2012. “Partisan asymmetries in online political activity.” *EPJ Data Science* 1(1): 1–19.

Conover, Michael et al. 2011. “Political Polarization on Twitter.” In *Proc. 5th Intl Conference on Weblogs and Social Media*,.

Conover, Michael, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, et al. 2011. “Predicting the political alignment of twitter users.” In *Privacy, Security, Risk and Trust(PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)*, IEEE, p. 192–199.

Conover, Michael et al. 2013. “The geospatial characteristics of a social movement communication network.” *PloS one* 8(3).

Coviello, Lorenzo et al. 2014. “Detecting emotional contagion in massive social networks.” *PloS one* 9(3): e90315.

DiGrazia, J., K. McKelvey, J. Bollen, and F. Rojas. 2013. “More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior.” *PLoS ONE* 8(11).

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. “Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter.” *PloS one* 6(12): e26752.

Eisenstein, Jacob, Brendan O’Connor, Noah A. Smith, and Eric P. Xing. 2014. “Diffusion of Lexical Change in Social Media.” *PLoS ONE* 9(11).

Fan, Rui, Jichang Zhao, Yan Chen, and Ke Xu. 2014. “Anger is more influential than joy: Sentiment correlation in Weibo.” *PLoS ONE*: e110184.

Farrell, Henry. 2012. “The Internet’s consequences for politics.” *Annual Review of Political Science* 15: 35–52.

Feller, Albert, Matthias Kuhnert, Timm O. Sprenger, and Isabell M. Welpe. 2011. “Divided They Tweet: The Network Structure of Political Microbloggers and Discussion Topics.” In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*,

Palo Alto, CA: Association for the Advancement of Artificial Intelligence (AAAI, p. 474–477.

Franch, Fabio. 2013. “Wisdom of the Crowds: 2010 UK Election Prediction with Social Media.” *in Journal of Information Technology and Politics* 10: 57–71.

Frank, Morgan R., Lewis Mitchell, Peter Sheridan Dodds, and Christopher M. Danforth. 2012. “Happiness and the patterns of life: a study of geolocated tweets.” *Scientific reports* 3: 2625–2625.

Gayo-Avello, Daniel. 2013. “A meta-analysis of state-of-the-art electoral prediction from Twitter data.” *Social Science Computer Review* 31(6): 649–679.

Ghiassi, M., J. Skinner, and D. Zimbra. 2013. “Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network.” *Expert Systems with applications* 40(16): 6266–6282.

Gleason, Benjamin. 2013. “#Occupy wall street: Exploring informal learning about a social movement on Twitter.” *American Behavioral Scientist*.

Gohdes, Anita R. 2015. “Pulling the plug Network disruptions and violence in civil conflict.” *Journal of Peace Research* 52(3): 352–367.

Golbeck, Jennifer, and Derek Hansen. 2014. “A method for computing political preference among Twitter followers.” *Social Networks* 36: 177–184.

Golder, Scott A., and Michael W. Macy. 2011. “Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures.” *Science* 333(6051): 1878–1881.

Golder, Scott, and Sarita Yardi. 2010. “Structural predictors of tie formation in twitter: Transitivity and mutuality.” *IEEE Second International Conference on Social Computing (SocialCom)* 2010: 88–95.

Golub, Benjamin, and Matthew O. Jackson. 2010. “Using selection bias to explain the observed structure of internet diffusions.” *Proceedings of the National Academy of Sciences* 107(24): 10833–10836.

Gomez, R., Leskovec J. Manuel, and A. Krause. 2010. “Inferring networks of diffusion and influence.” *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*

Gonzalez, Roberto, Ruben Cuevas, Angel Cuevas, and Carmen Guerrero. 2011. *Where are my followers? Understanding the Locality Effect in Twitter*. Arxiv.

Grossman, Guy, Macartan Humphreys, and Gabriella Sacramone-Lutz. 2014. “I wld like u WMP to extend electricity 2 our village: On Information Technology and Interest Articulation.” *American Political Science Review* 108(3): 688–705.

Hammond, Jesse, and Nils B Weidmann. 2014. “Using machine-coded event data for the micro-level study of political violence.” *Research & Politics* 1(2).

Hansen, Lars Kai, Adam Arvidsson, Finn Arup Nielsen, Elanor Colleoni, and Michael Etter. 2011. “Good friends, bad news-affect and virality in twitter.” In *Future information technology*, Berlin Heidelberg: Springer, p. 34–43.

Hawelka, Bartosz et al. 2014. “Geo-located Twitter as proxy for global mobility patterns.” *Cartography and Geographic Information Science* 41(3): 260–271.

Himelboim, Itai, Stephen McCreery, and Marc Smith. 2013. “Birds of a feather tweet together: Integrating network and content analyses to examine cross?ideology exposure on Twitter.” *Journal of Computer?Mediated Communication* 18(2): 40–60.

Hong, Liangjie, and Brian D. Davison. 2010. “Empirical study of topic modeling in twitter.” In *Proceedings of the First Workshop on Social Media Analytics*, ACM, p. 80–88.

Howard, Philip N., Sheetal D. Agarwal, and Muzammil M. Hussain. 2011. “When do states disconnect their digital networks? Regime responses to the political uses of social media.” *The Communication Review* 14(3): 216–232.

Howard, Philip N., and Muzammil M. Hussain. 2013. *Democracy's Fourth Wave?: Digital Media and the Arab Spring*. Oxford University Press.

Howard, Philip N., and Muzammil M. Hussain. 2011. “The upheavals in Egypt and Tunisia: the role of digital media.” *Journal of Democracy* 22(3): 35–48.

Huang, Shu, Wei Peng, Jingxuan Li, and Dongwon Lee. 2013. “Sentiment and topic analysis on social media: a multi-task multi-label classification approach.” In *Proceedings of the 5th annual ACM web science conference*, ACM, p. 172–181.

Hussain, Muzammil M., and Philip N. Howard. 2013. “What best explains successful protest cascades? ICTs and the fuzzy causes of the Arab Spring.” *International Studies Review* 15(1): 48–66.

Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. “Twitter power: Tweets as electronic word of mouth.” *Journal of the American society for information science and technology* 60(11): 2169–2188.

Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. 2007. “Why we twitter: understanding microblogging usage and communities.” In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, ACM, p. 56–65.

Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, et al. 2011. “Target-dependent twitter sentiment classification.” In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, Association for Computational Linguistics, p. 151–160.

Jungherr, Andreas, and Pascal Jurgens. 2013. "Forecasting the pulse: How deviations from regular patterns in online data can identify offline phenomena." *Internet Research* 23(5): 589–607.

Kalathil, Shanthi, and Taylor C. Boas. 2003. *Open Networks, Closed Regimes: The Impact of the Internet on Authoritarian Rule*. Washington, DC: Carnegie Endow.

Kaltenbrunner, A. et al. 2012. "Far from the eyes, close on the web: Impact of geographic distance on online social interactions." In *Proceedings of the 5th ACM Workshop on Online Social Networks (WOSN'12)*, Helsinki, Finland, p. 19–24.

Kim, Suin, JinYeong Bak, and Alice Haeyun Oh. 2012. "Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations." In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12)*,.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. "How censorship in China allows government criticism but silences collective expression." *American Political Science Review* 107(2): 326–343.

Kuehn, Christian, Erik A. Martens, and Daniel M. Romero. 2014. "Critical transitions in social network activity." *Journal of complex networks* 2(2): 141–152.

Kulshrestha, J., F. Kooti, A. Nikravesh, and Gummadi Kp. 2012. "Geographic dissection of the Twitter network." In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12)*, Dublin, Ireland, p. 202–209.

Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. 2010. "What is Twitter, a social network or a news media?" In *Proceedings of the 19th international conference on World wide web*, ACM, p. 591–600.

Lassen, David S., and Adam R. Brown. 2011. "Twitter: the electoral connection?" *Social Science Computer Review* 29(4): 419–436.

Lawrence, Eric, John Sides, and Henry Farrell. 2010. "Self-segregation or deliberation? Blog readership, participation, and polarization in American politics." *Perspectives on Politics* 8(01): 141–157.

Lazer, David et al. 2009. "Computational social science." *Science* 323(5915): 721–723.

Lee, Ryong, and Kazutoshi Sumiya. 2010. "Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection." In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks*, ACM, p. 1–10.

Leetaru, Kalev, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, et al. 2013. "Mapping the global Twitter heartbeat: The geography of Twitter." *First Monday* 18(5).

Lehmann, Janette, Bruno Gonçalves, José J. Ramasco, and Ciro Cattuto. 2012. "Dynamical classes of collective attention in twitter." In *Proceedings of the 21st international conference on World Wide Web*, ACM, p. 251–260.

Lerman, Kristina, and Rumi Ghosh. 2010. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." In *Proceeding of 4th international AAAI conference on weblogs and social media*, Washington, D. C., p. 90–97.

Lerman, Kristina, and Tad Hogg. 2010. "Using a model of social dynamics to predict popularity of news." In *Proceedings of the 19th international conference on World wide web*, ACM, p. 621–630.

Leskovec, Jure, Lada A. Adamic, and Bernardo A. Huberman. 2007. "The dynamics of viral marketing." *ACM Transactions on the Web (TWEB)* 1(1): 5.

Lin, Yu-Ru, Brian Keegan, Drew Margolin, and David Lazer. 2014. "Rising Tides or Rising Stars?: Dynamics of Shared Attention on Twitter during Media Events." *PLoS One* 9(5).

Lotan, Gilad, Erhardt Graeff, Mike Ananny, Devin Gaffney, et al. 2011. "The Arab Spring - the revolutions were tweeted: Information flows during the Tunisian and Egyptian revolutions." *International Journal of Communication* 5: 1375–405.

Lynch, Marc. 2011. "After Egypt: The limits and promise of online challenges to the authoritarian Arab state." *Perspectives on politics* 9(2): 301–310.

Martin-Shields, Charles, and Elizabeth Stones. 2014. "Smart Phones and Social Bonds: Communication Technology and Inter-Ethnic Cooperation in Kenya." *Journal of Peacebuilding & Development* 9(3): 50–64.

Mehrotra, Rishabh, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. "Improving lda topic models for microblogs via tweet pooling and automatic labeling." In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, ACM, p. 889–892.

Mei, Qiaozhu, Chao Liu, Hang Su, and ChengXiang Zhai. 2006. "A probabilistic approach to spatiotemporal theme pattern mining on weblogs." In *Proceedings of the 15th international conference on World Wide Web*, ACM, p. 533–542.

Metaxas, Panagiotis Takis, and Eni Mustafaraj. 2012. "Social media and elections." *Science* 338(6106): 472–473.

Metternich, Nils W., Cassy Dorff, Max Gallop, Simon Weschle, et al. 2013. "Antigovernment networks in civil conflicts: how network structures affect conflictual behavior." *American Journal of Political Science* 57(4): 892–911.

Metzger, Megan et al. 2014. *Dynamics of influence in online protest networks: Evidence from the 2013 Turkish protests*. Paper presented at the annual meeting of the Midwest Political Science Association.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, et al. 2011. "Understanding the Demographics of Twitter Users." In *5th International AAAI Conference on Weblogs and Social Media*,.

Mitchell, Lewis, Morgan R. Frank, Kameron Decker Harris, Peter Sheridan Dodds, and Christopher M. Danforth. 2013. "The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place." *PLoS ONE* 8(5).

Monroe, Burt L, Michael P Colaresi, and Kevin M Quinn. 2008. "Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict." *Political Analysis* 16(4): 372–403.

Morozov, Evgeny. 2011. "Whither Internet Control?" *Journal of Democracy* 22(2): 62–74.

Morris, Meredith Ringel, Scott Counts, Asta Roseway, Aaron Hoff, et al. 2012. "Tweeting is believing?: understanding microblog credibility perceptions." In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, p. 441–450.

Munger, Kevin. 2014. "Elites Tweet to get Feet off the Streets: Measuring Regime Response to Protest Using Social Media."

Mustafaraj, Eni, Samantha Finn, Carolyn Whitlock, and Panagiotis Takis Metaxas. 2011. "Vocal minority versus silent majority: discovering the opinions of the long tail." In *Proceedings of the 3rd IEEE International Conference on Social Computing*, Washington, D. C.

Mustafaraj, Eni, Samantha Finn, Carolyn Whitlock, and Panagiotis T. Metaxas. 2011. "Vocal minority versus silent majority: Discovering the opionions of the long tail." In *Privacy, Security, Risk and Trust (PASSAT) and IEEE Third Inernational Conference on Social Computing (SocialCom)*, , p. 103–110.

Mutz, Diana C. 2002. "Cross-cutting social networks: Testing democratic theory in practice." *American Political Science Review* 96(1): 111–126.

Naaman, Mor, Jeffrey Boase, and Chih-Hui Lai. 2010. "Is it really about me?: message content in social awareness streams." In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM, p. 189–192.

Naveed, Nasir, Thomas Gottron, Jérome Kunegis, and Arifah Che Alhadi. 2011. "Bad news travel fast: A content-based analysis of interestingness on twitter." In *Proceedings of the 3rd International Web Science Conference*, ACM, p. 8.

Nemeth, Stephen C, Jacob A Mauslein, and Craig Stapley. 2014. "The primacy of the local: Identifying terrorist hot spots using geographic information systems." *The Journal of Politics* 76(02): 304–317.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, , p. 1–2.

Pak, Alexander, and Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." *LREC* 10: 1320–1326.

Parmelee, John H., and Shannon L. Bichard. 2012. *Politics and the Twitter Revolution: How Tweets Influence the Relationship between Political Leaders and the Public*. Lanham, MD: Lexington Books.

Pfitzner, René, Antonios Garas, and Frank Schweitzer. 2012. "Emotional Divergence Influences Information Spreading in Twitter." In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, , p. 2–5.

Pierskalla, Jan H, and Florian M Hollenbach. 2013. "Technology and collective action: The effect of cell phone coverage on political violence in Africa." *American Political Science Review* 107(02): 207–224.

Prior, Markus. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. New York: Cambridge Univ. Press.

Ramakrishnan, Naren et al. 2014. "Beating the news' with EMBERS: forecasting civil unrest using open source indicators." In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, , p. 1799–1808.

Ringsquandl, Martin, and Dusan Petkovic. 2013. "Analyzing Political Sentiment on Twitter." In *AAAI Spring Symposium: Analyzing Microtext*, , p. 40–47.

Ritter, Daniel P., and Alexander H. Trechsel. 2014. "Revolutionary Cells: On the Role of Texts, Tweets, and Status Updates in Unarmed Revolutions." In *The Internet and Democracy in Global Perspective*, Springer International Publishing, p. 111–127.

Rød, Espen Geelmuyden, and Nils B Weidmann. 2015. "Empowering activists or autocrats? The Internet in authoritarian regimes." *Journal of Peace Research* 52(3): 338–351.

Romero, Daniel M., Brendan Meeder, and Jon Kleinberg. 2011. "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter." In *Proceedings of the 20th international conference on World wide web*, ACM, p. 695–704.

Sasahara, Kazutoshi, Yoshito Hirata, Masashi Toyoda, Masaru Kitsuregawa, et al. 2013. "Quantifying Collective Attention from Tweet Stream." *PLoS ONE* 8(4).

Schroeder, Rob, Sean F. Everton, and Russell Shepherd. 2014. "The Strength of Tweet Ties." In *Online Collective Action*, Vienna: Springer, p. 179–195.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and contagion are generically confounded in observational social network studies." *Sociological methods & research* 40(2): 211–239.

Shamma, David A., Lyndon Kennedy, and Elizabeth F. Churchill. 2011. "Peaks and persistence: modeling the shape of microblog conversations." In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, eds. Pamela Hinds et al. New York, NY: ACM, p. 355–358.

Shapiro, Jacob N, and Nils B Weidmann. 2015. "Is the Phone Mightier Than the Sword? Cellphones and Insurgent Violence in Iraq." *International Organization* 69(02): 247–274.

Shirky, Clay. 2011. "The political power of social media." *Foreign affairs* 90(1): 28–41.

Siegel, Alexandra. 2014. "Tweeting Beyond Tahrir: Ideological Diversity and Political Tolerance in Egyptian Twitter Networks."

Stieglitz, Stefan, and Linh Dang-Xuan. 2012. "Political Communication and Influence through Microblogging: An Empirical Analysis of Sentiment in Twitter Messages and Retweet Behavior." In *Proceedings of the 45th Hawaii International Conference on System Science*, ed. Ralph H. Sprague Jr. Washington, DC: IEEE Computer Society, p. 3500–3509.

Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed H. Chi. 2010. "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network." In *IEEE second international conference on Social computing (socialcom)*, , p. 177–184.

Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. 2012. "Geography of Twitter networks." *Social networks* 34(1): 73–81.

Thelwall, Mike, Kevan Buckley, and Georgios Paltoglou. 2011. "Sentiment in Twitter events." *Journal of the American Society for Information Science and Technology* 62(2): 406–418.

Theocharis, Yannis. 2013. "The wealth of (occupation) networks? Communication patterns and information distribution in a Twitter protest network." *Journal of Information Technology & Politics* 10(1): 35–56.

Tudoroiu, Theodor. 2014. "Social Media and Revolutionary Waves: The Case of the Arab Spring." *New Political Science* 36(3): 346–365.

Tufekci, Zeynep, and Christopher Wilson. 2012. "Social media and the decision to participate in political protest: Observations from Tahrir Square." *Journal of Communication* 62(2): 363–379.

Tumasjan, Andranik, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. 2010. "Predicting elections with twitter: What 140 characters reveal about political sentiment." In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, , p. 178–185.

Wang, Hao, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle." In *Proceedings of the ACL 2012 System Demonstrations*, Association for Computational Linguistics, p. 115–120.

Wang, Xiaofeng, Matthew S. Gerber, and Donald E. Brown. 2012. "Automatic crime prediction using events extracted from twitter posts." In *Social Computing, Behavioral-Cultural Modeling and Prediction*, Berlin Heidelberg: Springer, p. 231–238.

Ward, Michael D. et al. 2013. "Learning from the past and stepping into the future: Toward a new generation of conflict prediction." *International Studies Review* 15(4): 473–490.

Warren, T Camber. 2015. "Explosive connections? Mass media, social media, and the geography of collective violence in African states." *Journal of Peace Research*.

Warren, T. Camber. 2014. "Not by the sword alone: Soft power, mass media, and the production of state sovereignty." *International Organization* 68(01): 111–141.

Weidmann, Nils B. 2015. "Communication networks and the transnational spread of ethnic conflict." *Journal of Peace Research*.

Weng, Lillian, Alessandro Flammini, Alessandro Vespignani, and Fillipo Menczer. 2012. "Competition among memes in a world with limited attention." *Scientific reports* 2.

Windt, Peter Van der, and Macartan Humphreys. 2014. "Crowdseeding in Eastern Congo Using Cell Phones to Collect Conflict Events Data in Real Time." *Journal of Conflict Resolution*.

Wojcieszak, Magdalena E., and Diana C. Mutz. 2009. "Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?" *Journal of Communication* 59(1): 40–56.

Wolfsfeld, Gadi, Elad Segev, and Tamir Sheafer. 2013. "Social media and the Arab spring politics comes first." *The International Journal of Press/Politics* 18(2): 115–137.

Wu, Fang, and Bernardo A. Huberman. 2007. "Novelty and collective attention." *Proceedings of the National Academy of Sciences* 104(no. 45): 17599–17601.

Wu, Shaomei, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. "Who says what to whom on twitter." In *Proceedings of the 20th international conference on World wide web*, ACM, p. 705–714.

Yardi, Sarita, and Danah Boyd. 2010. "Dynamic debates: An analysis of group polarization over time on twitter." *Bulletin of Science, Technology & Society* 30(5): 316–327.

Yuan, Quan, Gao Cong, Zongyang Ma, Aixin Sun, et al. 2013. "Who, where, when and what: discover spatio-temporal topics for twitter users." In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, p. 605–613.

Zamal, Al, Wendy Liu Faiyaz, and Derek Ruths. 2012. "Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors." In *Proceedings of the International Conference on Weblogs and Social Media*,.

Zaman, Tauhid R., Ralf Herbrich, Jurgen Van Gael, and David Stern. 2010. "Predicting information spreading in twitter." *Workshop on computational social science and the wisdom of crowds* 104(45): 17599–601.

Zeitzoff, Thomas. 2013. "Conflict Dynamics, International Audiences, and Public Communication: Evidence from the 2012 Gaza Conflict." *Unpublished manuscript*.

Zeitzoff, Thomas, John Kelly, and Gilad Lotan. 2015. "Using social media to measure foreign policy dynamics An empirical analysis of the Iranian–Israeli confrontation (2012–13)." *Journal of Peace Research*.

Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. 2012. "Predicting asset value through twitter buzz." In *Advances in Collective Intelligence*, Berlin Heidelberg: Springer, p. 23–34.

Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. 2011. "Predicting stock market indicators through twitter "I hope it is not as bad as I fear." *Procedia-Social and Behavioral Sciences* 26: 55–62.

Appendix

Table A1. Nigerian Multilingual Dictionary of "armed conflict"

English	Arabic	French	Hausa	Igbo	Yoruba
aggression	عدوان	agression	ta'adi	awakpo	ifinran
aggressions	الاعتداءات	agressions	ta'addancin		
aggressor	معتدي	agresseur	tsokanar zalunci	ocho	
aggressors	المعتدين	agresseurs	tsõkana	ebido	
airstrike	غارة جوية	raid aérien	harin jirgin sama		
airstrikes	الضربات الجوية	frappes aériennes	harin na jiragen		
ak 47					
ak47					
ambush	كمين	embuscade	kwanto		
ambushed	كمين	embuscade	kwanton	echechiela	nibon
ambushes	الكمائن	embuscades	kwanton bauna	nēru nbì	ẹbu
ambushing	نصب الكمائن	embuscade			
annihilate	إبادة	annihiler	warware	ekpochapụ	
annihilated	يباد	anéanti	shafe	n'iyi	ọdi
annihilates	يقضي	annihile	shafe		
annihilating	هلك	annihilant	halakar	kpochapụ	
annihilation	بإبادة		rushewa	ebibi	
antagonism	تضاد	antagonisme	abotar gaba	imegidesi	
antagonist	خصم	antagoniste		na-eti ọkpọ	
antagonists	الخصوم	antagonistes		na-akụ ọkpọ	
armament	تسلح	armement	makamai	nke zaịọn	iham
armaments	التسلح	armements		ngwá agha	
armed	مسلح	armé		agha	ologun
armies	الجيوش	armées	sojojin	ụsụụ ndị agha	ogun
arming	تسليح	armement	tara makamai	igbochi ngwá agha ijupụta	
armored	مدرع	blindé	sulke		
armoured	مدرع	blindé	sulke		
army	جيش	armée	sojojin	agha	ogun
artillery	المدفعية	artillerie	manyan bindigogi	ogbunigwe	
assassinate	اغتيال	assassiner	kisa	igbu mmadu	
assassinated	اغتيال	assassiné	kashe	egbu	
assassinates	تغتال	assassine			
assassinating	اغتيال	assassinant	kisan gilla		
assassination	قتل	assassinat	kisan gilla	mgbu mmadu	
assassinations	الاغتيالات	assassinats	aikata kashe-kashen		ipania
assault	اعتداء	agression	hari	wakpo	sele si
assaulted	اعتداء	agressé	auka	tiri	nri ipalara
assaulting	الاعتداء	assaut		n'ịwakpo	
assaults	الاعتداءات	agressions	hari	ẹma ẹsịn	
attack	هجوم	attaque	hari	agha	kolu
attacked	هاجم	attaqué	sun kai hari	wakpoo	kolu
attacker	مهاجم	attaquant		ebibi	
attackers	المهاجمين	attaquants	maharan	kpara	
attacking	مهاجمة	attaquer	kai hare hare	awakpo	bàa
attacks	هجمات	attaques	kai hare-hare	ọgụ	ku
barricade	متراس		barikadi	mgbochi	
barricaded	تحصين	barricadé		mechibido	
barricades	المتاريس				
barricading	باعتراض	bastingages		imechibido	
battalion	كتيبة	bataillon	bataliya		ẹwú
battalions	كتائب	bataillons			ọrọrún
battle	معركة	bataille	yaƙi	agha	ogun
battled	تشابكت	lutté	fama	agha	
battlefield	ساحة المعركة	champ de bataille	fagen fama	n'ọgbọ agha	ogun

Table A1 (cont.) Nigerian Multilingual Dictionary of "armed conflict"

English	Arabic	French	Hausa	Igbo	Yoruba
battlefields	ساحات القتال	champs de bataille	fagen		
battlefront	جبهة القتال				
battlefronts	جبهات القتال	champs de bataille			
battleground	ساحة المعركة	champ de bataille	a fafata	agha	
battlegrounds	معارك	champs de bataille	dauki ba dadi		
battles	المعارك	batailles	fadace-fadace	agha	ogun
battleship	سفينة حربية	navire de guerre	jirgin ruwa na soja	agha	
battleships	البوارج	cuirassés			
battlespace	المعركة	bataille			
battlespaces		espaces de combat			
battling	تقاتل	combattre		alụ	njijadu
behead	قطع رأسه	décapiter			
beheaded	قطع رأس	décapité	fille kansa	isi	bẹ
beheading	قطع رأس	décapitation	fille		
beheadings	قطع الرؤوس	décapitations			
belligerent	دولة محاربة	belligérant		mmụọ ịlụ ọgụ	
belligerents	المتحاربين	belligérants			
bled	نزف	saigné	zub da jini		leemoô
bleed	ينزف	saigner	jinni	igba obara	
bleeding	نزيف	saignement	na jini	ọbara ọgbụgba	ẹjẹ
bleeds	ينزف	saigne			
blockade	حصار	blocus	kawancen	mgbochi	
blockaded	المحاصر	bloqué		nọchibidoro anọchibido	
blockades	الحصار	blocus			
blockading	حصار	blocus			
blood	دم	sang	jini	ọbara	ẹjẹ
bloodshed	سفك الدماء	effusion de sang	zubar da jini	na-awụfu ọbara	
bloodstain	بقع دم	tache de sang			
bloodstained	ملطخ بالدم	taché de sang		ọbara tetọrọ	
bloodstains	بقع الدم	taches de sang		ọbara	
bloody	دام	sanglant	na jini	ọbara	itajesile
bomb	قنبلة	bombe	bam	bọmbụ	bombu
bombed	قصف	bombardé	bamai	tụrụ bọmbụ	
bomber	مهاجم	bombardier		otu bombu	
bombers	المهجرين	bombardiers	kai harin	atụ bọmbụ	
bombing	قصف	bombardement	bom	bọmbụ	bombu
bombings	تفجيرات	attentats à la bombe	bom		
bombs	القنابل	bombes	ragargaza		ado-
brigade	لواء		birged	brigeedi	ẹgbẹ ọmọ ogun
brigades	ألوية				
bullet	رصاصة	balle	harsashi		ibọn
bullets	الرصاص	balles	harsasai	mgbọ	awako
casualties	خسائر	victimes	jikkata	onwu	faragbogbe
casualty	الإصابات	victime	mai hasara	a na-egbu	
combat	قتال		fama	ọgụ	ija
combatant	مقاتل	combattant		n'Ịlụ Agha	
combatants	المقاتلين	combattants		na-alụ agha	ogun
conflict	صراع	conflit	rikici	esemokwu	rogbodiyan
conflicts	الصراعات	conflits	rikice-rikice	esemokwu	ija
confrontation	مواجهة	affrontement	adawa		
confrontations	مواجهات			ese okwu	
coup	انقلاب		juyin mulki	kuu	
coups	الانقلابات		juyin mulki ne		
damage	ضرر	dommage		mmebi	bibajẹ
damaged	التالفة	endommagé	lalace	mebiri emebi	ti bajẹ
damaging	ضررا	dommageable	tareda žata	emebiri	ọmọde
dead	ميت	mort	matattu	nwụrụ anwụ	okú

Table A1 (cont.) Nigerian Multilingual Dictionary of "armed conflict"

English	Arabic	French	Hausa	Igbo	Yoruba
deadly	قاتل	mortel		na-egbu egbu	oloro
death	الموت	décès	mutuwa	ọnwụ	iku
deaths	وفيات	décès	mutuwar	ọnwụ	iku
decapitate	ضرب العنق	décapiter			
decapitated	مقطوعة الرأس	décapité			
decapitates	يقطع رأس	décapite			
decapitating	قطع رأس	décapitant			
decapitation	قطع الرأس	décapitation			
destroy	هدم	détruire	halaka	ebibi	
destroyed	دمرت	détruit	halakar	ebibi	
destroyer	مدمر	destructeur	hallakarwa	mbibi	apanirun
destroyers	مدمرات		hallaka	ebukọrọ	awọn afinișeiję
destroying	تدمير	détruisant	hallaka	ebibi	dabaru
destroys	يدمر	détruit	halaka	ebibie	
destruction	تدمير		halaka	mbibi	iparun
die	يموت	mourir	mutu	anwụ	kú
died	توفي	mort	ya rasu	nwụrụ	kú
dies	يموت	meurt	mutu	na-anwụ anwụ	ku
dismember	يمزق	démembrer			
dismembered	تقطيع اوصاله	démembré		emekwa	
dismembering	تمزيق	équarrissage			
dismembers	يقطع أوصال	démembre			
dying	الموت	mourant	mutuwa	na-anwụ anwụ	ku
enemies	الأعداء	ennemis	makiyan	iro	ọtá
enemy	العدو	ennemi	maƙiyi	onye iro	ọtá
explosion	انفجار		fashewa	gbawaranụ	bugbamu
explosive	مادة متفجرة	explosif		mgbawa	ibẹjadi
explosives	متفجرات	explosifs	nakiyoyi		
fatal	قاتلة			egbu egbu	apani
fatalities	وفيات	décès		anwụ	
fatality	نكبة	fatalité		ọdachi	
fatally	على نحو مهلك	mortellement		gbagburu	
feud	عداء	querelle	gaba	esemeokwu	orilede
feuded	احتدام	rivalisait			
feuding	المتناحرة	vendetta	husuma		mu awoôn
feuds	الحزازات	querelles			
fight	عراك	bats toi	yaki	agha	ija
fighter	مقاتل	combattant	jirgin saman soja		onija
fighters	مقاتل	combattants	mayakan	alụso	awọn onija
fighting	القتال	combat	fada	ọgụ	ija
fights	المعارك	combats	ta faɗa	ịlụ ọgụ	njà
firearm	سلاح ناري	arme à feu			ohun ija
firearms	الأسلحة النارية	armes à feu	bindigogi	eji égbè agbagbu	ibon
firefight	معركة	fusillade			
firefights	معارك	des échanges de tirs			
force	قوة		karfi	ike	agbara
forces	القوات		sojojin	agha	ologun
fought	قاتل	combattu	suka yi jihãdi	agha	ja
grave	قبر	tombe	kabari	ili	sin
graves	القبور	tombes	kaburbura	ili	ibojì
grenade	قنبلة يدوية		gurnati	bombu	
grenades	قنابل		gurnetin		
guerillas	حزب	guérilleros	dakarun	agha okpuru	
guerrilla	حرب العصابات	guérilla	yaƙin	okpuru	
gun	بندقية	pistolet	bindiga	egbe	ibon
gunboat	زورق حربي	canonnière			
gunboats	الزوارق الحربية	canonnières			

Table A1 (cont.) Nigerian Multilingual Dictionary of “armed conflict”

English	Arabic	French	Hausa	Igbo	Yoruba
gunfire	إطلاق نار	des coups de feu	bindigar		
gunman	مسلح	tireur			
gunmen	مسلحون	des hommes armés	yan bindiga		
gunned	قتل	abattu			
gunner	مدفعي	canonnier	sojan igwa	onye agha	
gunners	المدفعية	canonniers			
gunning	علم المسدسات				
gunpowder	بارود	poudre à canon			
guns	البنادق	pistolets	bindigogi	egbe	ibon
gunship	حربية				
gunships	طائرات	hélicoptères de combat			
gunshot	إطلاق نار	coup de feu	harbin bindiga		ìbọn
gunshots	طلق ناري	des coups de feu	bindigogi	ụda égbè	
handgun	مسدس	pistolet			
handguns	المسدسات	armes de poing		égbè mkpụmkpụ	
hostiles	معادية				
hostilities	الأعمال العدائية	hostilités	tashin		igboro
hostility	عداء	hostilité	rashin jituwa	iro	igbogunti
infantry	المشاة	infanterie	dakaru	bipu	ẹlẹsẹ
ied	العبوات الناسفة		bama-bamai		
ieds	العبوات الناسفة		bamai		
injure	جرح	blesser	cuta	emerụ	ipalara
injured	جرح	blessé	ji rauni	merụrụ ahụ	farapa
injures	إصابات	blesse		emerụ	
injuries	إصابات	blessures	raunin da ya faru	unan	nosi
injuring	إصابة	blessant	jikkata	merụọ	
injury	ضرر	blessure	rauni	mmerụ	ipalara
insurgencies	التمرد	insurrections	hare haren		
insurgency	تمرد	insurrection	tayar da kayar baya		
insurgent	متمرد	insurgé	hare		
insurgents	المتمردين	insurgés	maharan		
invade	غزو	envahir	mamaye	wakporo	gbogun
invaded	غزت	envahi	mamaye	wakporo	yabo
invader	غاز	envahisseur	mai mamaye	onye mbusoagha	
invaders	الغزاة	envahisseurs		mwakpo	
invades	يغزو	envahit	ta mamaye	awakpoo	
invading	الغازية	envahisseur		na-awakwasị	
invasion	غزو		mamayewa	mbuso agha	ayabo
invasions	الغزوات		mamayar	mwakpo	
kill	قتل	tuer	kashe	igbu	pa
killed	قتل	tué	kashe	gburu	pa
killer	القاتل	tueur	kisa	egbu egbu	apani
killers	القتلة	tueurs	kisan		aporó
killing	قتل	meurtre	kashe	okowot	pipa
killings	القتل	tueries	kashe-kashe		
kills	يقتل	tue	kashe	egbu	pa
land mine	لغم أرضي	mine terrestre		ala m	ilẹ mi
land mines	الألغام الأرضية	les mines terrestres	ƙasar mahakai	ogbunigwe	ilẹ maini
landmine	الألغام الأرضية	les mines terrestres			
landmines	الألغام الأرضية	les mines terrestres	nakiyoyin da		lese
machinegun	رشاش	mitraillette			
machineguns	الرشاشة	mitrailleuses			
maim	جرح	mutiler			maimu
maimed	شوهوا	estropié		nkwarụ	àbùkù
maiming	التشويه	mutilation			sofo ti
maims	يشوه	mutile			

Table A1 (cont.) Nigerian Multilingual Dictionary of “armed conflict”

English	Arabic	French	Hausa	Igbo	Yoruba
marines	المارينز		sojin rundunar jiragen ruwa		marini
massacre	مذبحة		kisan kiyashin	mgbuchapụ	ipakupa
massacred	ذبح	massacrés	karkashe		
massacres	المجازر		kisan kiyashi		
massacring	ذبح	massacrant			
militarize	أضفى الصفة العسكرية	militariser	sojoji		
militarized	عسكرتها	militarisée	yan bindiga a	agha	
militarizing	عسكرة	militarisation			
military	عسكري	militaire	soja	agha	ologun
missile	صاروخ		makami mai linzami	agha	misaili
missiles	صواريخ		jifa	akụ ụta	
mortar	هاون	mortier	turmi	ngwa agha	amọ
mortars	قذائف الهاون	mortiers			
murder	قتل	assassiner	kisankai	igbu ọchụ	iku
murdered	قتل	assassiné	kashe	gburu	paniyan
murderer	قاتل	assassin	kisan kai	na-egbu ọchụ	apànìyàn
murderers	القتلة	meurtriers	kisankai	na-egbu ọchụ	apànìyàn
murdering	اغتيال	meurtre	kashe	-egbu ọchụ,	
murderous	القاتل	meurtrier	suka kai	igbu ọchụ	ipaniyan
murderously	مهلك				
murders	القتل	meurtres	kisan kai	ikwa	
mutilate	بتر	mutiler	daddatsa gawa	ebepụ	
mutilated	المشوهة	mutilé			
mutilates	يمزق	mutile			
mutilating	تشويه	mutilant	jikin		
naval	بحري		sojan ruwa		to ọkọ
navies	القوات البحرية	marines			
navy	سلاح البحرية		sojojin ruwa	agha mmiri	ọgagun
ordinance	مرسوم	ordonnance	farilla	ukpuru	ìlana
ordinances	المراسيم	ordonnances	hukuncen		idajọ
pistol	مسدس	pistolet	bindiga	egbe	ibon
pistols	مسدسات	pistolets		obere égbè	
platoon	فرزة		mutanena su ka		
platoons	فصائل	pelotons			
raid	غارة		hari	wakporo	igbogun ti
raided	داهمت	perquisitionné	kai hari	wabara	
raiding	الإغارة	raids	hari		ẹgbẹ ogun
raids	الغارات		hare-hare		
rape	اغتصاب	viol	fyade	n'ike	ifipabanilopo
raped	اغتصاب	violé	fyade	n'ike	lopọ ti
rapes	الاغتصاب	viols			ifipabanilopo
raping	اغتصاب	viol	fyade		
rapist	مغتصب	violeur	yarsu fyaden		
rapists	المغتصبين	violeurs			afipabanilo
rebel	متمرد	rebelle	yan tawayen	enupu isi	ṣọtẹ
rebelled	تمرد	rebellés	tawaye	nupụrụ isi	ṣọtẹ
rebelling	تمرد	rebeller	tawaye	enupụ isi	ti ṣọtẹ
rebellion	تمرد	rébellion	tawayen	nnupụisi	iṣọtẹ
rebellions	التمرد	rébellions	tawayen	nupụ isi	
rebellious	متمرد	rebelle		enupụ isi	ọlọtẹ
rebels	الثوار	rebelles	yan tawayen	nnupụisi	olote
revolt	ثورة	révolte	yi tawaye	nnupụisi	sote
revolts	الثورات	révoltes	yin tawaye	nnupụisi	
revolver	مسدس			égbè	
revolvers	المسدسات				
rifle	بندقية	fusil	bindiga	égbè	ibọn

Table A1 (cont.) Nigerian Multilingual Dictionary of "armed conflict"

English	Arabic	French	Hausa	Igbo	Yoruba
rifleman	حامل بندقية	fusilier			
riflemen	الرماه	tirailleurs			
rifles	بنادق	fusils	bindigogi		awọn iru ibọn kan
riot	شغب	émeute		ntịme	ìṣọtẹ na
rioted	قاموا بأعمال الشغب	se sont révoltés			
rioter	متظاهر	émeutier			
rioters	مثيري الشغب	émeutiers	masu zanga-zangar		
rioting	أعمال شغب	émeutes			
riots	أعمال الشغب	émeutes	tarzoma		
rocket	صاروخ	fusée	roka	rọketi	
rocketfire	نيران الصواريخ				
rocketlauncher	قاذفة الصواريخ	lance-roquettes			
rocketlaunchers					
rockets	صواريخ	roquettes	roka	tammy	
security	أمن	sécurité	tsaro	nche	aabo
shelled	قشر	décortiquées			
shelling	قصف	bombardement	wuta ya janyo		
shotgun	بندقية الصيد	fusil de chasse			ibọn
shotguns	البنادق	fusils de chasse			
slaughter	ذبح	abattage	kashe	akwụ	
slaughtered	ذبح	abattus	yanka	gbuo	pa
slaughtering	ذبح	abattage	yanka	ogbugbu	ẹran
slaughters	مجازر	tueries	yanka		
small arms	الأسلحة الصغيرة	petites armes	kananan makamai	obere ogwe aka	kekere apá
sniper	قناص	tireur isolé	maharbi		
snipers	القناصة		makasa		orukoô
soldier	جندي	soldat	soja	agha	jagunjagun
soldiers	جنود	soldats	sojoji	agha	ogun
stabbed	طعن	poignardé	sukan	adụ	leyiti
stabbing	طعن	élancement	caka	ịma	nibi
strike	إضراب	grève	yajin	iku	idasesile
strikes	الضربات	grèves	buga	etiwapụ	dasofo
striking	ضرب عن العمل	frappant	daukan hankali	pụtara ìhè	idaṣẹ
struck	ضرب	frappé	bugi	gburu	lù
suicidal	الانتحار	suicidaire		igbu onwe	
suicide	انتحار		kashe kansa	igbu onwe	ara
terror	الإرهاب	terreur	tsõro	oké ụjọ	eruolorun
terrorise	إرهاب	terroriser	ta'ada	menyeujo	
terrorised	روعت	terrorisé			
terrorises	يرهب	terrorise			
terrorising	إرهاب	terrorisant	tayar da hankalin		
terrorism	إرهاب	terrorisme	ta'addanci	iyi ọha egwu	ipanilaya
terrorist	إرهابي	terroriste	'yan ta'adda	eyi ọha egwu	apanilaya
terrorists	الإرهابيين	terroristes	'yan ta'adda	eyi ọha egwu	onijagidijagan
terrorize	إرهاب	terroriser	ta'ada	menyeujo	
terrorized	روعت	terrorisé			
terrorizes	ترعب	terrorise	barazana,		
terrorizing	إرهاب	terrorisant	tayar da hankalin		
threat	التهديد	menace	barazana	iyi egwu	irokeke
threaten	هدد	menacer	barazana	ize	deruba
threatened	مهددة	menacés	barazana	egwu	ewu
threatening	مهدد	menaçant	barazana	na-eyi egwu	ihal
threateningly	مهددا	menaçant		egwu	
threatens	يهدد	menace	barazana	egwu	irokeke
threats	التهديدات	menaces	barazana	egwu	irokeke

Table A1 (cont.) Nigerian Multilingual Dictionary of “armed conflict”

English	Arabic	French	Hausa	Igbo	Yoruba
troop	قوات	troupe	ƙungiya		
troops	القوات	troupes	dakarun	agha	enia
victim	ضحية	victime	wanda aka azabtar	aja	njiya
victims	ضحايا	victimes	wadanda ke fama	metụtara	olufaragba
violence	عنف		tashin hankali	ime ihe ike	iwa-ipa
violent	عنيف			eme ihe ike	iwa
violently	بعنف	violemment		ike	
war	حرب	guerre	yaki	agha	ogun
warfare	حرب	guerre	yaki	agha	yce
warfighter	المحاربون	combattant			
warfighters	قوات قتالة	combattants			
warmonger	مثير الحرب	belliciste			
warmongers	دعاة الحرب	bellicistes			
warplane	طائرة حربية	avion de combat			
warplanes	طائرات حربية	avions de combat	jirage		
warred	حارب	guerroyé	yaki	agha	n gbogun ti
warring	قتال	en guerre	yaƙe	ebu agha	
warrior	محارب	guerrier			jagunjagun
warriors	المحاربين	guerriers		dike	
wars	الحروب	guerres	yaƙe-yaƙe	agha	ogun
warship	سفينة حربية	navire de guerre			
warships	سفن حربية	navires de guerre			
wartorn	مزقته الحرب				
weapon	سلاح	arme	makami	ngwá agha	multani
weaponry	أسلحة	armes	makamai	ngwá agha	
weapons	أسلحة	armes	makamai	ngwá agha	ohun ija
wound	جرح	blessure	rauni	ọnya	egbo
wounded	جريح	blessés	rauni	merụrụ	ti o gbọgbẹ
wounding	جرح	blessant	ji masa rauni		
wounds	الجروح	plaies	raunuka	ọnyá	ọgbẹ